

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using 8w model

Run on: December 3, 2003, 15:40:35 ; Search time 8.3653 Seconds

(without alignments)  
171.565 Million cell updates/sec

Title: US-09-788-308D-1

Perfect score: 41

Sequence: 1 XXPHLKRK 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A-Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	85.4	955	22	ABR71635
2	34	82.9	170	22	AA870494
3	33	80.5	6	17	AA87271
4	33	80.5	7	17	AA87274
5	33	80.5	9	22	AAU07650
6	33	80.5	19	16	AA875440
7	33	80.5	19	17	AA87279
8	33	80.5	22	16	AA875444
9	33	80.5	22	17	AA87278

10	33	80.5	22	22	AAU07653	Human surfactant p
11	33	80.5	22	22	AAU07654	Human surfactant p
12	33	80.5	22	22	AAU07655	Human surfactant p
13	33	80.5	22	22	AAU07656	Human surfactant p
14	33	80.5	22	22	AAU07657	Human surfactant p
15	33	80.5	22	22	AAU07658	Human surfactant p
16	33	80.5	22	22	AAU07659	Human surfactant p
17	33	80.5	22	22	AAU07660	Human surfactant p
18	33	80.5	22	22	AAU07661	Human surfactant p
19	33	80.5	23	16	AA875439	Synthetic peptide
20	33	80.5	23	16	AA875441	Synthetic peptide
21	33	80.5	23	16	AA875443	Synthetic peptide
22	33	80.5	23	16	AA875445	Synthetic peptide
23	33	80.5	23	16	AA875446	Synthetic peptide
24	33	80.5	23	16	AA875447	Synthetic peptide
25	33	80.5	23	16	AA875448	Synthetic peptide
26	33	80.5	27	14	AA844388	Surfactant apoprot
27	33	80.5	27	14	AA844389	Surfactant apoprot
28	33	80.5	27	14	AA844394	Surfactant apoprot
29	33	80.5	27	14	AA844395	Surfactant apoprot
30	33	80.5	27	14	AA844396	Surfactant apoprot
31	33	80.5	27	14	AA844397	Surfactant apoprot
32	33	80.5	27	15	AA862811	Hydrophobic, lung
33	33	80.5	27	15	AA862812	Hydrophobic, lung
34	33	80.5	27	15	AA862817	Hydrophobic, lung
35	33	80.5	27	15	AA862818	Hydrophobic, lung
36	33	80.5	27	15	AA862819	Hydrophobic, lung
37	33	80.5	27	15	AA862820	Hydrophobic, lung
38	33	80.5	27	16	AA875437	Synthetic peptide
39	33	80.5	27	16	AA875438	Synthetic peptide
40	33	80.5	27	16	AA875439	Synthetic peptide
41	33	80.5	27	16	AA875441	Synthetic peptide
42	33	80.5	27	16	AA875442	Synthetic peptide
43	33	80.5	28	9	AA80579	Human SAR(Val) pep
44	33	80.5	32	16	AA875435	Synthetic peptide
45	33	80.5	34	22	AA851583	Amino acid fragment

#### ALIGNMENTS

RESULT 1	ABR71635	standard; Protein; 955 AA.
ID	ABR71635	
XX	ABR71635	
DT	26-MAR-2002 (first entry)	
XX		
DE	Drosophila melanogaster polypeptide SEQ ID NO 41697.	
XX		
KW	Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.	
KW		
XX		
OS	Drosophila melanogaster.	
XX		
PN	WO200171042-A2.	
XX		
PD	27-SEP-2001.	
XX		
PF	23-MAR-2001; 2001WO-US09231.	
XX		
PR	23-MAR-2000; 2000US-191637P.	
XX		
PR	11-JUL-2000; 2000US-0614150.	
XX		
PA	(PEKE ) PE CORP NY.	
XX		
PI	Venter JC, Adams M, Li PWD, Myers EW,	
XX		
DR	WPI: 2001-656860/75.	
XX		
XX	N-PSDB; ABL15738.	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more	



AC AAR97274;  
XX  
XX 11-FEB-1997 (first entry)  
DT  
XX  
DE Intermediate for prodn. of surfactant peptide.  
XX  
XX Surfactant; respiratory distress syndrome; intermediate; soluble;  
KM treatment.  
XX  
XX Synthetic.  
OS  
FH Key Location/Qualifiers  
FT Modified-site 1 /note= "Fmoc-Cys(Acm)"  
FT Modified-site 4 /note= "His(Trt)"  
FT Modified-site 6 /note= "Lys(Boc)"  
FT Modified-site 7 /note= "Arg(Mcr)"  
FT  
XX  
XX WO9617872-A1.  
XX 13-JUN-1996.  
XX  
XX 06-JUN-1995; 95WO-JP01114.  
XX  
XX 07-DEC-1994; 94JP-0303397.  
XX  
XX (TANB ) TOKYO TANABE CO.  
XX  
XX Ohtsubo E, Takei T;  
XX  
XX WPI; 1996-287121/29.  
XX  
XX Peptide intermediate for production of surfactant peptide(s) - used  
PT in lung surfactants for treatment of respiratory distress syndrome  
XX  
XX Claim 2; Page 14; 19pp; Japanese.  
XX  
XX The present sequence is an intermediate for the prodn. of  
CC a surfactant peptide, which may be incorporated into lung  
CC surfactant formulations for the treatment of respiratory distress  
CC syndrome. The surfactant peptide prep. using the present  
CC sequence is highly soluble in, e.g. methanol, is readily  
CC compounded with lipid mixes. and has good suspendability and  
CC surfactant activity.  
XX  
XX SQ Sequence 7 AA;  
XX  
XX Query Match 80.5%; Score 33; DB 17; Length 7;  
XX Best Local Similarity 100.0%; Pred. No. 9.3e+05; Indels 0; Gaps 0;  
XX Matches 6; Conservative 0; Mismatches 0;  
QY 3 PVHLKR 8  
DB 2 PVHLKR 7  
XX  
XX RESULT 5  
XX AAU07650 standard; Peptide; 9 AA.  
XX  
XX AAU07650;  
XX  
XX 04-DEC-2001 (first entry)  
XX  
XX Human surfactant protein spreading agent generic peptoid sequence.  
DE  
XX Human; surfactant protein B; surfactant protein C; SP-B; SP-C; peptoid;  
KM pulmonary; protein therapy; spreading agent; N-substituted glycine;  
KM lung surfactant; pulmonary surfactant; alveolar surface activity;  
KM respiratory distress syndrome.

XX  
XX Homo sapiens.  
OS Synthetic.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 1..2 /label= Phe, OTHER  
FT /note= "OTHER= palmitoylated cysteine,  
FT phenylmethyl glycine, octylamine glycine or  
FT hexadecylamine glycine"  
FT  
FT Modified-site 9 /note= "asb or spe substituted"  
FT  
XX  
XX WO200160837-A2.  
XX  
XX 23-AUG-2001.  
XX  
XX 16-FEB-2001; 2001WO-US05145.  
XX  
XX 16-FEB-2000; 2000US-0182847.  
XX  
XX (NOUN ) UNIV NORTHWESTERN.  
XX (CHIR ) CHIRON CORP.  
XX  
XX Barron AE, Zuckermann RN, Wu CW;  
XX  
XX WPI; 2001-550045/61.  
XX  
XX Heteropolymeric pulmonary spreading agent having at least one  
PT N-substituted glycine residue and an amino acid residue corresponding  
PT to a natural surfactant-associated protein, useful for treating lung  
PT respiratory distress -  
XX  
XX Claim 28; Page 27; 40pp; English.  
XX  
XX The invention relates to a non-natural heteropolymeric pulmonary  
CC spreading agent comprising at least one N-substituted glycine residue and  
CC at least one amino acid residue corresponding to the surfactant proteins  
CC B and C (SP-B and SP-C). These surfactant protein mimics are peptoid  
CC sequences (reverse sequence of the natural protein i.e. equal to the  
CC carboxy to amino sequence of the peptide) added to a lipid mixture to  
CC create a functional, non-immunogenic lung surfactant with physiological  
CC alveolar surface activity. The peptoid sequences of the invention can be  
CC used to enhance the solubility of surfactant associated proteins (to  
CC therefore enhance resistance to aggregation) and can also affect alveolar  
CC surface tension during an inhalation/exhalation cycle. The spreading  
CC agents are useful for treating disorders of the lungs such as respiratory  
CC distress syndrome. This sequence, represents a human surfactant protein  
CC spreading agent generic peptide sequence.  
XX  
XX SQ Sequence 9 AA;  
XX  
XX Query Match 80.5%; Score 33; DB 22; Length 9;  
XX Best Local Similarity 100.0%; Pred. No. 9.13e+05; Indels 0; Gaps 0;  
XX Matches 6; Conservative 0; Mismatches 0;  
QY 3 PVHLKR 8  
DB 3 PVHLKR 8  
XX  
XX RESULT 6  
XX AAR75440 standard; peptide; 19 AA.  
XX  
XX AAR75440;  
XX  
XX 01-FEB-1996 (first entry)  
XX  
XX Synthetic peptide used in a lung surfactant.  
DE  
XX Lung surfactant; respiratory distress syndrome; hydrophobic; peptide.  
KM  
KM

```
OS Synthetic.
XX
XX MO9515980-A1.
XX
XX 15-JUN-1995.
XX
XX 07-DEC-1994; 94WO-JP02057.
XX
XX 08-DEC-1993; 93JP-0307657.
XX
XX (TANAB ) TOKYO TANABE CO.
XX
XX Ohkawa H, Ohtsubo E, Takei T;
XX
XX WPI, 1995-224289/29.
XX
XX Lung surfactant containing new synthetic peptide - having an
XX hydrophobic C-terminal chain, for treatment of respiratory distress
XX syndrome
XX
XX Claim 5; Page 43; 54pp; Japanese.
XX
XX AAR75432-R75452 are synthetic peptides with hydrophobic C-terminal
XX chains. The peptides can be formulated with a lipid mixture
XX (choleline phosphoglyceride, acidic phospholipid and a fatty acid)
XX to give a lung surfactant. The surfactant is used to treat
XX respiratory stress disorder.
XX
XX Sequence 19 AA;
XX
XX Query Match 80.5%; Score 33; DB 16; Length 19;
XX Best Local Similarity 100.0%; Pred. No. 4.8;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 3 PVLKLR 8
XX |||||
XX 2 PVLKLR 7
XX
XX Db
XX
XX RESULT 7
XX AAR97279
XX ID AAR97279 standard; peptide; 19 AA.
XX
XX AC AAR97279;
XX
XX DT 11-FEB-1997 (first entry)
XX
XX DE Surfactant peptide.
XX
XX KM Surfactant; respiratory distress syndrome; intermediate; soluble;
XX treatment.
XX
XX OS Synthetic.
XX
XX PN WO9617872-A1.
XX
XX PD 13-JUN-1996.
XX
XX PF 06-JUN-1995; 95WO-JP01114.
XX
XX PR 07-DEC-1994; 94JP-0303397.
XX
XX PA (TANAB ) TOKYO TANABE CO.
XX
XX PI Ohtsubo E, Takei T;
XX
XX DR WPI, 1996-287121/29.
XX
XX Peptide intermediate for production of surfactant peptide(s) - used
XX in lung surfactants for treatment of respiratory distress syndrome
XX
XX Example 2; Page 13; 19pp; Japanese.
XX
```

```
CC The present sequence is a surfactant peptide (SP) prepd. from the
CC SP intermediate of the invention, which may be incorporated into
CC lung surfactant formulations for the treatment of respiratory
CC distress syndrome. The present peptide is highly soluble in, e.g.
CC methanol, is readily compounded with lipid mixes. and has good
CC suspendability and surfactant activity. A surfactant powder contg.
CC the present peptide, when tested using the method of WO9321225,
CC lowers surface tension from 34.1 to 3.9 dyne/cm.
XX
XX Sequence 19 AA;
XX
XX Query Match 80.5%; Score 33; DB 17; Length 19;
XX Best Local Similarity 100.0%; Pred. No. 4.8;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 3 PVLKLR 8
XX |||||
XX 2 PVLKLR 7
XX
XX Db
XX
XX RESULT 8
XX AAR75444
XX ID AAR75444 standard; peptide; 22 AA.
XX
XX AC AAR75444;
XX
XX DT 01-FEB-1996 (first entry)
XX
XX DE Synthetic peptide used in a lung surfactant.
XX
XX KM Lung surfactant; respiratory distress syndrome; hydrophobic; peptide.
XX
XX OS Synthetic.
XX
XX PH Key Location/Qualifiers
XX FT Misc-difference 7..22
XX FT /label= N1e
XX FT /note= "amino acids 7 to 22 are all Norleucines"
XX
XX PN WO9515980-A1.
XX
XX PD 15-JUN-1995.
XX
XX PF 07-DEC-1994; 94WO-JP02057.
XX
XX PR 08-DEC-1993; 93JP-0307657.
XX
XX PA (TANAB ) TOKYO TANABE CO.
XX
XX PI Ohkawa H, Ohtsubo E, Takei T;
XX
XX DR WPI, 1995-224289/29.
XX
XX PT Lung surfactant containing new synthetic peptide - having an
XX hydrophobic C-terminal chain, for treatment of respiratory distress
XX syndrome
XX
XX PS Claim 5; Page 45; 54pp; Japanese.
XX
XX CC AAR75432-R75452 are synthetic peptides with hydrophobic C-terminal
XX chains. The peptides can be formulated with a lipid mixture
XX (choleline phosphoglyceride, acidic phospholipid and a fatty acid)
XX to give a lung surfactant. The surfactant is used to treat
XX respiratory stress disorder.
XX
XX SQ Sequence 22 AA;
XX
XX Query Match 80.5%; Score 33; DB 16; Length 22;
XX Best Local Similarity 100.0%; Pred. No. 5.6;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 3 PVLKLR 8
XX |||||
```



Db 1 PVLKR 6

## RESULT 9

AA97278 standard; peptide; 22 AA.

AA97278;

11-FEB-1997 (first entry)

Surfactant peptide.

Surfactant; respiratory distress syndrome; intermediate; soluble; treatment.

Synthetic.

Key Location/Qualifiers

Modified-site 7 /label= N1e

Modified-site 8 /label= N1e

Modified-site 9 /label= N1e

Modified-site 10 /label= N1e

Modified-site 11 /label= N1e

Modified-site 12 /label= N1e

Modified-site 13 /label= N1e

Modified-site 14 /label= N1e

Modified-site 15 /label= N1e

WO9617872-A1.

13-JUN-1996.

06-JUN-1995; 95WO-0P01114.

07-DEC-1994; 94JP-0303397.

(TANB ) TOKYO TANABE CO.

Ohtsubo E, Takei T;

WPI; 1996-287121/29.

Peptide intermediate for production of surfactant peptide(s) - used in lung surfactants for treatment of respiratory distress syndrome

Example 2; Page 12; 19pp; Japanese.

The present sequence is a surfactant peptide (SP) prepd. from the SP intermediate of the invention, which may be incorporated into lung surfactant formulations for the treatment of respiratory distress syndrome. The present peptide is highly soluble in, e.g. methanol, is readily compounded with lipid mixts. and has good suspendability and surfactant activity. A surfactant powder contg. the present peptide, when tested using the method of WO9321225, lowers surface tension from 34.1 to 3.9 dyne/cm.

Sequence 22 AA;

Query Match 80.5%; Score 33; DB 17; Length 22;

Best Local Similarity 100.0%; Pred. No. 5.6; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVLKR 8

Db 1 PVLKR 6

## RESULT 10

AA07653 standard; Peptide; 22 AA.

AA07653;

04-DEC-2001 (first entry)

Human surfactant protein C aromatic peptoid-peptide mimic SPCM.

Human; surfactant protein B; surfactant protein C; SP-B, SP-C; peptoid; pulmonary; protein therapy; spreading agent; N-substituted glycine;

lung surfactant; pulmonary surfactant; alveolar surface activity; respiratory distress syndrome.

Homo sapiens.

Synthetic.

Key Location/Qualifiers

Modified-site 1..2 /note= "Hexadecylamine glycine"

Modified-site 9 /note= "Phenylmethyl glycine"

Modified-site 10..11 /note= "spe substituents"

Modified-site 12 /note= "spe substituents"

Modified-site 13..14 /note= "phenylmethyl glycine"

Modified-site 15 /note= "spe substituents"

Modified-site 16..17 /note= "phenylmethyl glycine"

Modified-site 18 /note= "spe substituents"

Modified-site 19..20 /note= "phenylmethyl glycine"

Modified-site 21 /note= "spe substituents"

Modified-site 22 /note= "spe substituent"

Modified-site 22 /note= "spe substituent"

Modified-site 22 /note= "C-terminal amide"

WO200160837-A2.

23-AUG-2001.

16-FEB-2001; 2001WO-US05145.

16-FEB-2000; 2000US-0182847.

(NOUN ) UNIV NOIRON CORP.

Barron AE, Zuckermann RN, Wu CW;

WPI; 2001-550045/61.

Heteropolymetric pulmonary spreading agent having at least one N-substituted glycine residue and an amino acid residue corresponding to a natural surfactant-associated protein, useful for treating lung respiratory distress -

Claim 20; Fig 7A; 40pp; English.

The invention relates to a non-natural heteropolymetric pulmonary spreading agent comprising at least one N-substituted glycine residue and at least one amino acid residue corresponding to the surfactant proteins B and C (SP-B and SP-C). These surfactant protein mimics are peptoid

CC sequences (reverse sequence of the natural protein i.e. equal to the  
CC carboxy to amino sequence of the peptide) added to a lipid mixture to  
CC create a functional, non-immunogenic, lung surfactant with physiological  
CC alveolar surface activity. The peptid sequences of the invention can be  
CC used to enhance the solubility of surfactant associated proteins (to  
CC therefore enhance resistance to aggregation) and can also affect alveolar  
CC surface tension during an inhalation/exhalation cycle. The spreading  
CC agents are useful for treating disorders of the lungs such as respiratory  
CC distress syndrome. This sequence represents a human surfactant protein C  
CC mimic peptid.

CC  
XX  
SQ Sequence 22 AA;

Query Match 80.5%; Score 33; DB 22; Length 22;  
Best Local Similarity 100.0%; Pred. No. 5.6;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 PVLKLR 8  
3 PVLKLR 8  
3 PVLKLR 8

Db

RESULT 11  
AAU07654  
ID AAU07654 standard; Peptide; 22 AA.  
XX  
AC AAU07654;  
XX  
DT 04-DEC-2001 (first entry)  
XX  
DE Human surfactant protein C aromatic peptid-peptide mimic SPCM2.  
XX  
KW Human; surfactant protein B; surfactant protein C; SP-B; SP-C; peptid;  
KW pulmonary; protein therapy; spreading agent; N-substituted glycine;  
KW lung surfactant; pulmonary surfactant; alveolar surface activity;  
KW respiratory distress syndrome.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 1..2  
FT Modified-site /note= "Octylamine glycine"  
FT Modified-site 9  
FT Modified-site /note= "Phenylmethyl glycine"  
FT Modified-site 10..11  
FT Modified-site /note= "spe substituents"  
FT Modified-site 12  
FT Modified-site /note= "Phenylmethyl glycine"  
FT Modified-site 13..14  
FT Modified-site /note= "spe substituents"  
FT Modified-site 15  
FT Modified-site /note= "Phenylmethyl glycine"  
FT Modified-site 16..17  
FT Modified-site /note= "spe substituents"  
FT Modified-site 18  
FT Modified-site /note= "Phenylmethyl glycine"  
FT Modified-site 19..20  
FT Modified-site /note= "spe substituents"  
FT Modified-site 21  
FT Modified-site /note= "spe substituent"  
FT Modified-site 22  
FT Modified-site /note= "spe substituent"  
FT Modified-site 22  
FT Modified-site /note= "C-terminal amide"  
XX  
XX  
PN WO200160837-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 16-FEB-2001; 2001WO-US05145.  
XX  
PR 16-FEB-2000; 2000US-0182847.

XX  
PA (NOUN ) UNIV NOIRON CORP.  
XX  
PI Barron AE, Zuckermann RN, Wu CW;  
XX  
XX WPI; 2001-550045/61.  
DR  
XX  
XX Heteropolymetric pulmonary spreading agent having at least one  
PT N-substituted glycine residue and an amino acid residue corresponding  
PT to a natural surfactant-associated protein, useful for treating lung  
PT respiratory distress -  
XX  
XX  
PS Claim 20; Fig 7A; 40pp; English.  
XX  
XX  
CC The invention relates to a non-natural heteropolymetric pulmonary  
CC spreading agent comprising at least one N-substituted glycine residue and  
CC at least one amino acid residue corresponding to the surfactant proteins  
CC B and C (SP-B and SP-C). These surfactant protein mimics are peptid  
CC sequences (reverse sequence of the natural protein i.e. equal to the  
CC carboxy to amino sequence of the peptide) added to a lipid mixture to  
CC create a functional, non-immunogenic lung surfactant with physiological  
CC alveolar surface activity. The peptid sequences of the invention can be  
CC used to enhance the solubility of surfactant associated proteins (to  
CC therefore enhance resistance to aggregation) and can also affect alveolar  
CC surface tension during an inhalation/exhalation cycle. The spreading  
CC agents are useful for treating disorders of the lungs such as respiratory  
CC distress syndrome. This sequence represents a human surfactant protein C  
CC mimic peptid.

XX  
SQ Sequence 22 AA;

Query Match 80.5%; Score 33; DB 22; Length 22;  
Best Local Similarity 100.0%; Pred. No. 5.6;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 PVLKLR 8  
3 PVLKLR 8  
3 PVLKLR 8

Db

RESULT 12  
AAU07655  
ID AAU07655 standard; Peptide; 22 AA.  
XX  
AC AAU07655;  
XX  
DT 04-DEC-2001 (first entry)  
XX  
DE Human surfactant protein C aromatic peptid-peptide mimic SPCM3.  
XX  
KW Human; surfactant protein B; surfactant protein C; SP-B; SP-C; peptid;  
KW pulmonary; protein therapy; spreading agent; N-substituted glycine;  
KW lung surfactant; pulmonary surfactant; alveolar surface activity;  
KW respiratory distress syndrome.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 9  
FT Modified-site /note= "Phenylmethyl glycine"  
FT Modified-site 10..11  
FT Modified-site /note= "spe substituents"  
FT Modified-site 12  
FT Modified-site /note= "Phenylmethyl glycine"  
FT Modified-site 13..14  
FT Modified-site /note= "spe substituents"  
FT Modified-site 15  
FT Modified-site /note= "Phenylmethyl glycine"  
FT Modified-site 16..17  
FT Modified-site /note= "spe substituents"  
FT Modified-site 18  
FT Modified-site /note= "Phenylmethyl glycine"

```

FT Modified-site 19..20 /note= "spe substituents"
PT Modified-site 21 /note= "spe substituent"
FT Modified-site 22 /note= "spe substituent"
FT Modified-site 22 /note= "spe substituent"
FT Modified-site 22 /note= "C-terminal amide"
PN IIO200160837-A2.
PD 23-AUG-2001.
XX 16-FEB-2001; 2001WO-US05145.
XX 16-FEB-2000; 2000US-0182847.
XX (NOUN ) UNIV NOIRON CORP.
PA Barron AE, Zuckermann RN, Wu CW;
PI WPI; 2001-550045/61.
DR Heteropolymeric pulmonary spreading agent having at least one
XX N-substituted glycine residue and an amino acid residue corresponding
PT to a natural surfactant-associated protein, useful for treating lung
PT respiratory distress -
PS Claim 20; Fig 7A; 40pp; English.
XX The invention relates to a non-natural heteropolymeric pulmonary
CC spreading agent comprising at least one N-substituted glycine residue and
CC at least one amino acid residue corresponding to the surfactant proteins
CC B and C (SP-B and SP-C). These surfactant protein mimics are peptid
CC sequences (reverse sequence of the natural protein i.e. equal to the
CC carboxy to amino sequence of the peptide) added to a lipid mixture to
CC create a functional, non-immunogenic, lung surfactant with physiological
CC alveolar surface activity. The peptid sequences of the invention can be
CC used to enhance the solubility of surfactant associated proteins (to
CC therefore enhance resistance to aggregation) and can also affect alveolar
CC surface tension during an inhalation/exhalation cycle. The spreading
CC agents are useful for treating disorders of the lungs such as respiratory
CC distress syndrome. This sequence represents a human surfactant protein C
CC mimic peptoid.
XX Sequence 22 AA:
SQ
Query Match 80.5%; Score 33; DB 22; Length 22;
Best Local Similarity 100.0%; Pred. No. 5,6;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 3 PVHLKR 8
| | | | |
Db 3 PVHLKR 8
RESULT 13
AAU07656
ID AAU07656 standard; Peptide; 22 AA.
AC AAU07656;
XX
XX 04-DEC-2001 (first entry)
DT Human surfactant protein C aliphatic peptoid-peptide mimic SPCM4.
DE
XX Human; surfactant protein B; surfactant protein C; SP-B; SP-C; peptoid;
KW pulmonary; protein therapy; spreading agent; N-substituted glycine;
KW lung surfactant; pulmonary surfactant; alveolar surface activity;
XX respiratory distress syndrome.
XX Homo sapiens
OS Synthetic.
```

XX		Key	Location/Qualifiers 1..2
FH		Modified-site	/note= "Hexadecylamine glycine"
FT		Modified-site	9..22
FT		Modified-site	/note= "asb substituents"
FT		Modified-site	22
FT			/note= "C-terminal amide"
PN		WO200160837-A2.	
PN		23-Aug-2001.	
PD		16-FEB-2001; 2001WO-US05145.	
PF		16-FEB-2000; 2000US-0182847.	
PR		(NOUN ) UNIV NOIRON CORP.	
XX		Barron AE, Zuckermann RN, Wu CW;	
PI		WPI; 2001-550045/61.	
DR		Heteropolymeric pulmonary spreading agent having at least one	
XX		N-substituted glycine residue and an amino acid residue corresponding	
PT		to a natural surfactant-associated protein, useful for treating lung	
PT		respiratory distress -	
XX		Claim 20; Fig 7B; 40pp; English.	
PS		The invention relates to a non-natural heteropolymeric pulmonary	
CC		spreading agent comprising at least one N-substituted glycine residue and	
CC		at least one amino acid residue corresponding to the surfactant proteins	
CC		C B and C (SP-B and SP-C). These surfactant protein mimics are peptid	
CC		sequences (reverse sequence of the natural protein i.e. equal to the	
CC		cathoxy to amino sequence of the peptide) added to a lipid mixture to	
CC		create a functional, non-immunogenic lung surfactant with physiological	
CC		alveolar surface activity. The peptid sequences of the invention can be	
CC		used to enhance the solubility of surfactant associated proteins (to	
CC		therefore enhance resistance to aggregation) and can also affect alveolar	
CC		surface tension during an inhalation/exhalation cycle. The spreading	
CC		agents are useful for treating disorders of the lungs such as respiratory	
CC		distress syndrome. This sequence represents a human surfactant protein C	
CC		mimic peptid.	
SQ		Sequence 22 AA:	
OY	3 PVHLKR 8	Query Match	80.5%; Score 33; DB 22; Length 22;
Dd	3 PVHLKR 8	Best Local Similarity	100.0%; Pred. No. 5.6;
		Matches 6; Conservative 0; Mismatches	0; Indels 0; Gaps 0;
RESULT 14			
AU07657		ID AU07657 standard; Peptide; 22 AA.	
AAU07657;			
AC		04-DEC-2001 (first entry)	
DT			
XX			
XX			
DE		Human surfactant protein C aliphatic peptid-peptide mimic SPCMS.	
KM		Human; surfactant protein B; surfactant protein C; SP-B; SP-C; peptoid;	
KM		pulmonary; proteain therapy; spreading agent; N-substituted glycine;	
KM		lung surfactant; pulmonary surfactant; alveolar surface activity;	
KM		respiratory distress syndrome.	
OS		Homo sapiens.	
OS		Synthetic.	

```
XX Key Location/Qualifiers
FH Modified-site 1..2
FT Modified-site /note= "Oocylamine glycine"
FT Modified-site 9..22
FT Modified-site /note= "ssb substituents"
FT Modified-site 22
FT /note= "C-terminal amide"
XX
XX MO200160837-A2.
XX
XX 23-AUG-2001.
XX
XX 16-FEB-2001; 2001WO-US05145.
XX
XX 16-FEB-2000; 2000US-0182847.
XX
XX (NOUN ) UNIV NOIRON CORP.
XX
XX Barron AE, Zuckermann RN, Wu CW;
XX
XX WPI, 2001-550045/61.
XX
XX Heteropolymetric pulmonary spreading agent having at least one
XX N-substituted glycine residue and an amino acid residue corresponding
XX to a natural surfactant-associated protein, useful for treating lung
XX respiratory distress -
XX
XX PS Claim 20; Fig 7B; 40pp; English.
XX
XX The invention relates to a non-natural heteropolymetric pulmonary
XX spreading agent comprising at least one N-substituted glycine residue and
XX at least one amino acid residue corresponding to the surfactant proteins
XX B and C (SP-B and SP-C). These surfactant protein mimics are peptoid
XX sequences (reverse sequence of the natural protein i.e. equal to the
XX carboxy to amino sequence of the peptide) added to a lipid mixture to
XX create a functional, non-immunogenic lung surfactant with physiological
XX alveolar surface activity. The peptoid sequences of the invention can be
XX used to enhance the solubility of surfactant associated proteins (to
XX therefore enhance resistance to aggregation) and can also affect alveolar
XX surface tension during an inhalation/exhalation cycle. The spreading
XX agents are useful for treating disorders of the lungs such as respiratory
XX distress syndrome. This sequence represents a human surfactant protein C
XX mimic peptoid.
XX
XX SQ Sequence 22 AA;
XX
XX Query Match 80.5%; Score 33; DB 22; Length 22;
XX Best Local Similarity 100.0%; Pred. No. 5.6;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 3 PVHLKR 8
XX |||||
XX 3 PVHLKR 8
XX
XX DB
XX
XX RESULT 15
XX AAU07658
XX ID AAU07658 standard; Peptide; 22 AA.
XX
XX AC AAU07658;
XX
XX 04-DEC-2001 (first entry)
XX
XX Human surfactant protein C aliphatic peptoid-peptide mimic SPCM6.
XX
XX Human; surfactant protein B; surfactant protein C; SP-B; SP-C; peptoid;
XX pulmonary; protein therapy; spreading agent; N-substituted glycine;
XX lung surfactant; pulmonary surfactant; alveolar surface activity;
XX respiratory distress syndrome.
XX
XX OS Homo sapiens.
XX
XX OS Synthetic.
```

```
XX Key Location/Qualifiers
FH Modified-site 9..22
FT Modified-site /note= "ssb substituents"
FT Modified-site 22
FT /note= "C-terminal amide"
XX
XX MO200160837-A2.
XX
XX 23-AUG-2001.
XX
XX 16-FEB-2001; 2001WO-US05145.
XX
XX 16-FEB-2000; 2000US-0182847.
XX
XX (NOUN ) UNIV NOIRON CORP.
XX
XX Barron AE, Zuckermann RN, Wu CW;
XX
XX WPI, 2001-550045/61.
XX
XX Heteropolymetric pulmonary spreading agent having at least one
XX N-substituted glycine residue and an amino acid residue corresponding
XX to a natural surfactant-associated protein, useful for treating lung
XX respiratory distress -
XX
XX PS Claim 20; Fig 7B; 40pp; English.
XX
XX The invention relates to a non-natural heteropolymetric pulmonary
XX spreading agent comprising at least one N-substituted glycine residue and
XX at least one amino acid residue corresponding to the surfactant proteins
XX B and C (SP-B and SP-C). These surfactant protein mimics are peptoid
XX sequences (reverse sequence of the natural protein i.e. equal to the
XX carboxy to amino sequence of the peptide) added to a lipid mixture to
XX create a functional, non-immunogenic lung surfactant with physiological
XX alveolar surface activity. The peptoid sequences of the invention can be
XX used to enhance the solubility of surfactant associated proteins (to
XX therefore enhance resistance to aggregation) and can also affect alveolar
XX surface tension during an inhalation/exhalation cycle. The spreading
XX agents are useful for treating disorders of the lungs such as respiratory
XX distress syndrome. This sequence represents a human surfactant protein C
XX mimic peptoid.
XX
XX SQ Sequence 22 AA;
XX
XX Query Match 80.5%; Score 33; DB 22; Length 22;
XX Best Local Similarity 100.0%; Pred. No. 5.6;
XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 3 PVHLKR 8
XX |||||
XX 3 PVHLKR 8
XX
XX DB
XX
XX Search completed: December 3, 2003, 15:49:52
XX Job time : 8.32653 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 3, 2003, 15:47:35 ; Search time 2.87755 Seconds  
(Without alignments)  
132.334 Million cell updates/sec

Title: US-09-788-308d-1

Sequence: 1 XXPHLXRG 9

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued, Patents AA: \*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCOTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	87.8	284	4	US-09-252-991A-22908
2	35	85.4	140	4	US-09-252-991A-27630
3	33	80.5	12	6	5223481-2
4	33	80.5	12	6	5223481-3
5	33	80.5	12	6	5223481-4
6	33	80.5	12	6	5455227-1
7	33	80.5	12	6	5455227-3
8	33	80.5	12	6	5455227-6
9	33	80.5	17	6	5223481-8
10	33	80.5	17	6	5455227-4
11	33	80.5	19	2	US-08-652-450A-9
12	33	80.5	22	2	US-08-652-450A-13
13	33	80.5	23	2	US-08-652-450A-8
14	33	80.5	23	2	US-08-652-450A-10
15	33	80.5	23	2	US-08-652-450A-12
16	33	80.5	23	2	US-08-652-450A-14
17	33	80.5	23	2	US-08-652-450A-15
18	33	80.5	23	2	US-08-652-450A-16
19	33	80.5	23	2	US-08-652-450A-17
20	33	80.5	27	1	US-08-343-427B-1
21	33	80.5	27	1	US-08-343-427B-2
22	33	80.5	27	1	US-08-343-427B-7
23	33	80.5	27	1	US-08-343-427B-8
24	33	80.5	27	1	US-08-343-427B-9
25	33	80.5	27	1	US-08-343-427B-10
26	33	80.5	27	2	US-08-652-450A-6
27	33	80.5	27	2	US-08-652-450A-7

28	33	80.5	27	2	US-08-652-450A-11	Sequence 11, Appl
29	33	80.5	27	2	US-08-652-450A-20	Sequence 20, Appl
30	33	80.5	27	2	US-08-652-450A-21	Sequence 21, Appl
31	33	80.5	32	2	US-08-652-450A-4	Sequence 4, Appl
32	33	80.5	35	1	US-08-343-427B-11	Sequence 11, Appl
33	33	80.5	35	2	US-08-652-450A-1	Sequence 1, Appl
34	33	80.5	35	2	US-08-652-450A-5	Sequence 5, Appl
35	33	80.5	35	2	US-08-750-194-1	Sequence 1, Appl
36	33	80.5	35	6	5223481-5	Patent No. 5223481
37	33	80.5	35	6	5223481-6	Patent No. 5223481
38	33	80.5	35	6	5223481-7	Patent No. 5223481
39	33	80.5	35	6	5455227-2	Patent No. 5455227
40	33	80.5	186	2	US-08-750-194-2	Sequence 2, Appl
41	33	80.5	723	4	US-09-849-334-2	Sequence 2, Appl
42	32	78.0	207	4	US-09-252-991A-16678	Sequence 16678, A
43	32	78.0	222	4	US-09-252-991A-30831	Sequence 30831, A
44	32	78.0	594	4	US-09-252-991A-20849	Sequence 20849, A
45	31	75.6	139	3	US-09-120-663-4	Sequence 4, Appl

## ALIGNMENTS

```
RESULT 1
US-09-252-991A-22908
; Sequence 22908, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22908
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-22908
```

Query Match 87.8%; Score 36; DB 4; Length 284;  
Best Local Similarity 85.7%; Pred. No. 3.9;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 PVHLXRG 9  
Db 203 PVHLXRG 209

```
RESULT 2
US-09-252-991A-27630
; Sequence 27630, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27630
; LENGTH: 140
; TYPE: PRT
```

```
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27630

Query Match
Best Local Similarity 85.4%; Score 35; DB 4; Length 140;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKRG 9
   |||||
Db 45 PVHLORG 51

RESULT 3
5223481-2
; Patent No. 5223481
; APPLICANT: CURSTEDT, TORE;JORNVAL, HANS;LOWENADLER, BJORN;
; ROBERTSSON, BENGT
; TITLE OF INVENTION: BIOLOGICALLY ACTIVE LIPOPROTEIN AND
; ITS USE
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/423,346
; FILING DATE: 18-OCT-1989
; SEQ ID NO:2
; LENGTH: 12

5223481-2
Query Match
Best Local Similarity 80.5%; Score 33; DB 6; Length 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKR 8
   |||||
Db 7 PVHLKR 12

RESULT 4
5223481-3
; Patent No. 5223481
; APPLICANT: CURSTEDT, TORE;JORNVAL, HANS;LOWENADLER, BJORN;
; ROBERTSSON, BENGT
; TITLE OF INVENTION: BIOLOGICALLY ACTIVE LIPOPROTEIN AND
; ITS USE
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/423,346
; FILING DATE: 18-OCT-1989
; SEQ ID NO:3
; LENGTH: 12

5223481-3
Query Match
Best Local Similarity 80.5%; Score 33; DB 6; Length 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKR 8
   |||||
Db 7 PVHLKR 12

RESULT 5
5223481-4
; Patent No. 5223481
; APPLICANT: CURSTEDT, TORE;JORNVAL, HANS;LOWENADLER, BJORN;
; ROBERTSSON, BENGT
; TITLE OF INVENTION: BIOLOGICALLY ACTIVE LIPOPROTEIN AND
; ITS USE
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/423,346
; FILING DATE: 18-OCT-1989
; SEQ ID NO:4
; LENGTH: 12
```

```
5223481-4
Query Match
Best Local Similarity 80.5%; Score 33; DB 6; Length 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKR 8
   |||||
Db 7 PVHLKR 12

RESULT 6
5455227-1
; Patent No. 5455227
; APPLICANT: CURSTEDT, TORE;LOWENADLER, BJORN;ROBERTSON, BENGT
; TITLE OF INVENTION: BIOLOGICALLY ACTIVE LIPOPROTEIN AND ITS USE
; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/64,382
; FILING DATE: 21-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 423,346
; FILING DATE: 18-OCT-1989
; SEQ ID NO:1
; LENGTH: 12

5455227-1
Query Match
Best Local Similarity 80.5%; Score 33; DB 6; Length 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKR 8
   |||||
Db 7 PVHLKR 12

RESULT 7
5455227-3
; Patent No. 5455227
; APPLICANT: CURSTEDT, TORE;LOWENADLER, BJORN;ROBERTSON, BENGT
; TITLE OF INVENTION: BIOLOGICALLY ACTIVE LIPOPROTEIN AND ITS USE
; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/64,382
; FILING DATE: 21-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 423,346
; FILING DATE: 18-OCT-1989
; SEQ ID NO:3
; LENGTH: 12

5455227-3
Query Match
Best Local Similarity 80.5%; Score 33; DB 6; Length 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKR 8
   |||||
Db 7 PVHLKR 12

RESULT 8
5455227-6
; Patent No. 5455227
; APPLICANT: CURSTEDT, TORE;LOWENADLER, BJORN;ROBERTSON, BENGT
; TITLE OF INVENTION: BIOLOGICALLY ACTIVE LIPOPROTEIN AND ITS USE
; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/64,382
; FILING DATE: 21-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 423,346
; FILING DATE: 18-OCT-1989
```

SEQ ID NO:6;  
LENGTH: 12

Query Match 80.5%; Score 33; DB 6; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.6;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVLKR 8  
| | | | |  
DB 7 PVLKR 12

RESULT 9  
5223481-8  
Patent No. 5223481  
APPLICANT: CURSTEDT, TORE;JOENVALL, HANS;LOMENADLER, BJORN;  
ROBERTSSON, BENGT  
TITLE OF INVENTION: BIOLOGICALLY ACTIVE LIPOPROTEIN AND  
ITS USE  
NUMBER OF SEQUENCES: 8  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/423,346  
FILING DATE: 18-OCT-1989  
SEQ ID NO:8:  
LENGTH: 17  
5223481-8

Query Match 80.5%; Score 33; DB 6; Length 17;  
Best Local Similarity 100.0%; Pred. No. 0.87;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVLKR 8  
| | | | |  
DB 7 PVLKR 12

RESULT 10  
5455227-4  
Patent No. 5455227  
APPLICANT: CURSTEDT, TORE;LOMENADLER, BJORN;ROBERTSON, BENGT  
TITLE OF INVENTION: BIOLOGICALLY ACTIVE LIPOPROTEIN AND ITS USE  
NUMBER OF SEQUENCES: 6  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/64,382  
FILING DATE: 21-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 423,346  
FILING DATE: 18-OCT-1989  
SEQ ID NO:4:  
LENGTH: 17  
5455227-4

Query Match 80.5%; Score 33; DB 6; Length 17;  
Best Local Similarity 100.0%; Pred. No. 0.87;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVLKR 8  
| | | | |  
DB 7 PVLKR 12

RESULT 11  
US-08-652-450A-9  
Sequence 9, Application US/08652450A  
Patent No. 5827825  
GENERAL INFORMATION:  
APPLICANT: TAKEI, TSUNETOMO  
APPLICANT: OHTSUBO, EIJI  
APPLICANT: OHKAWA, HIROSHI  
TITLE OF INVENTION: NOVEL SYNTHETIC PEPTIDE, LUNG SURFACTANT  
TITLE OF INVENTION: CONTAINING THE SAME AND REMEDY FOR RESPIRATORY DISTRESS  
SYNDROME

NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GRIFFIN, BUTLER, WHISENUNT & KURTOSKY  
STREET: 2300 SOUTH NINTH STREET, SUITE PH-1  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: U.S.A.  
ZIP: 22204

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/652,450A  
FILING DATE: 05-JUN-1996  
CLASSIFICATION: 514

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 307657/1993  
FILING DATE: 08-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SZIEPL, JOERG-UWE

REGISTRATION NUMBER: 31,799  
REFERENCE/DOCKET NUMBER: A08A0006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 979-5700  
TELEFAX: (703) 979-7429

INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHEICAL: NO  
FRAGMENT TYPE: N-terminal  
US-08-652-450A-9

Query Match 80.5%; Score 33; DB 2; Length 19;  
Best Local Similarity 100.0%; Pred. No. 0.98;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVLKR 8  
| | | | |  
DB 2 PVLKR 7

RESULT 12  
US-08-652-450A-13  
Sequence 13, Application US/08652450A  
Patent No. 5827825  
GENERAL INFORMATION:  
APPLICANT: TAKEI, TSUNETOMO  
APPLICANT: OHTSUBO, EIJI  
APPLICANT: OHKAWA, HIROSHI  
TITLE OF INVENTION: NOVEL SYNTHETIC PEPTIDE, LUNG SURFACTANT  
TITLE OF INVENTION: CONTAINING THE SAME AND REMEDY FOR RESPIRATORY DISTRESS  
SYNDROME  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GRIFFIN, BUTLER, WHISENUNT & KURTOSKY  
STREET: 2300 SOUTH NINTH STREET, SUITE PH-1  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: U.S.A.  
ZIP: 22204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/652,450A

FILING DATE: 05-JUN-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 307657/1993  
FILING DATE: 08-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SZIPL, JOERG-UWE  
REGISTRATION NUMBER: 31,799  
REFERENCE/DOCKET NUMBER: AOB40006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 979-5700  
TELEFAX: (703) 979-7429  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: N-terminal  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 7..22  
OTHER INFORMATION: /product= "No. 5827825leucine"  
US-08-652-450A-13

Query Match 80.5%; Score 33; DB 2; Length 22;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 PVHLKR 8  
Db 1 PVHLKR 6

RESULT 13  
US-08-652-450A-8  
Sequence 8, Application US/08652450A  
Patent No. 5827825  
GENERAL INFORMATION:  
APPLICANT: TAKEI, TSUNETOMO  
APPLICANT: OHTSUBO, EIJI  
APPLICANT: OKAWA, HIROSHI  
TITLE OF INVENTION: NOVEL SYNTHETIC PEPTIDE, LUNG SURFACTANT  
TITLE OF INVENTION: CONTAINING THE SAME AND REMEDY FOR RESPIRATORY DISTRESS  
TITLE OF INVENTION: SYNDROME  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GRIFFIN, BUTLER, WISEHUNT & KURTOSKY  
STREET: 2300 SOUTH NINTH STREET, SUITE PH-1  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: U.S.A.  
ZIP: 22204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/652,450A  
FILING DATE: 05-JUN-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 307657/1993  
FILING DATE: 08-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SZIPL, JOERG-UWE  
REGISTRATION NUMBER: 31,799  
REFERENCE/DOCKET NUMBER: AOB40006  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 979-5700  
TELEFAX: (703) 979-7429  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: N-terminal  
US-08-652-450A-8

Query Match 80.5%; Score 33; DB 2; Length 23;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 PVHLKR 8  
Db 2 PVHLKR 7

RESULT 14  
US-08-652-450A-10  
Sequence 10, Application US/08652450A  
Patent No. 5827825  
GENERAL INFORMATION:  
APPLICANT: TAKEI, TSUNETOMO  
APPLICANT: OHTSUBO, EIJI  
APPLICANT: OKAWA, HIROSHI  
TITLE OF INVENTION: NOVEL SYNTHETIC PEPTIDE, LUNG SURFACTANT  
TITLE OF INVENTION: CONTAINING THE SAME AND REMEDY FOR RESPIRATORY DISTRESS  
TITLE OF INVENTION: SYNDROME  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GRIFFIN, BUTLER, WISEHUNT & KURTOSKY  
STREET: 2300 SOUTH NINTH STREET, SUITE PH-1  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: U.S.A.  
ZIP: 22204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/652,450A  
FILING DATE: 05-JUN-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 307657/1993  
FILING DATE: 08-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SZIPL, JOERG-UWE  
REGISTRATION NUMBER: 31,799  
REFERENCE/DOCKET NUMBER: AOB40006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 979-5700  
TELEFAX: (703) 979-7429  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: N-terminal  
US-08-652-450A-10

Query Match 80.5%; Score 33; DB 2; Length 23;  
Best Local Similarity 100.0%; Pred. No. 1.2;



Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 PVLKLR 8  
| | | | |  
Db 2 PVLKLR 7

RESULT 15  
US-08-652-450A-12  
; Sequence 12, Application US/08652450A  
; Patent No. 5827825  
; GENERAL INFORMATION:  
; APPLICANT: TAKEI, TSUNETOMO  
; APPLICANT: OHTSUBO, EIJI  
; APPLICANT: OHKAWA, HIROSHI  
; TITLE OF INVENTION: NOVEL SYNTHETIC PEPTIDE, LUNG SURFACTANT  
; TITLE OF INVENTION: CONTAINING THE SAME AND REMEDY FOR RESPIRATORY DISTRESS  
; TITLE OF INVENTION: SYNDROME  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GRIFFIN, BUTLER, WHISENHUNT & KURTOSKY  
; STREET: 2300 SOUTH NINTH STREET, SUITE PH-1  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: U.S.A.  
; ZIP: 22204  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/652,450A  
; FILING DATE: 05-JUN-1996  
; CLASSIFICATION: 514  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: JP 307657/1993  
; FILING DATE: 08-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SZIPL, JOERG-UWE  
; REGISTRATION NUMBER: 31,799  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 979-5700  
; TELEFAX: (703) 979-7429  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 23 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: N-terminal  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 8..23  
; OTHER INFORMATION: /product= "No. 5827825leucine"  
; OTHER INFORMATION: /label= Xaa  
US-08-652-450A-12

Query Match 80.5%; Score 33; DB 2; Length 23;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVLKLR 8  
| | | | |  
Db 2 PVLKLR 7

Search completed: December 3, 2003, 15:54:25  
Job time : 2.87755 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 15:50:00 ; Search time 5.44898 Seconds  
(Without alignments)  
307.186 Million cell updates/sec

Title: US-09-788-308D-1  
Perfect score: 41  
Sequence: 1 XXPVHLKRG 9

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubppa/PTCTUS\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppa/PTCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppa/US10\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubppa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	95.1	9	11	US-09-788-308D-1
2	39	95.1	24	11	US-09-788-308D-4
3	33	80.5	35	11	US-09-788-308D-3
4	33	80.5	44	9	US-09-864-761-41348
5	33	80.5	92	10	US-09-984-245-278
6	33	80.5	92	10	US-09-966-262-278
7	33	80.5	92	11	US-09-983-966-278
8	33	80.5	92	15	US-10-143-090-278
9	33	80.5	197	14	US-10-074-247-2
10	33	80.5	216	9	US-09-925-302-482
11	33	80.5	257	9	US-09-925-301-1016
12	33	80.5	257	15	US-10-023-896-59
13	33	80.5	257	15	US-10-023-896-87
14	33	80.5	713	15	US-10-116-519-8
15	33	80.5	723	15	US-10-274-878-2

16	32	78.0	137	9	US-09-864-761-42650	Sequence 42650, A
17	32	78.0	5179	9	US-09-922-217-1068	Sequence 1068, Ap
18	32	78.0	5179	10	US-09-833-263-1068	Sequence 1068, Ap
19	32	78.0	5179	14	US-10-025-380-1068	Sequence 1068, Ap
20	31	75.6	139	9	US-09-879-248-10	Sequence 10, Appl
21	31	75.6	643	15	US-10-274-878-4	Sequence 71, Appl
22	31	75.6	643	15	US-10-239-607-71	Sequence 138, App
23	30	73.2	112	12	US-10-078-113-188	Sequence 138, App
24	30	73.2	112	15	US-10-090-182-118	Sequence 1167, App
25	30	73.2	143	10	US-09-764-868-1167	Sequence 297, App
26	30	73.2	485	16	US-10-080-170-297	Sequence 61, Appl
27	30	73.2	943	12	US-10-148-351-61	Sequence 34248, A
28	30	73.2	2665	9	US-09-864-761-34248	Sequence 423, App
29	30	73.2	3664	15	US-10-177-293-423	Sequence 268, App
30	29	70.7	33	8	US-08-424-550B-268	Sequence 9, Appl
31	29	70.7	111	12	US-10-083-446-9	Sequence 10, Appl
32	29	70.7	111	12	US-10-083-446-10	Sequence 11, Appl
33	29	70.7	111	12	US-10-083-446-11	Sequence 12, Appl
34	29	70.7	111	12	US-10-083-446-12	Sequence 13, Appl
35	29	70.7	111	12	US-10-083-446-13	Sequence 14, Appl
36	29	70.7	111	12	US-10-083-446-14	Sequence 17, Appl
37	29	70.7	111	12	US-10-083-446-15	Sequence 18, Appl
38	29	70.7	111	12	US-10-083-446-16	Sequence 22, Appl
39	29	70.7	111	12	US-10-083-446-17	Sequence 23, Appl
40	29	70.7	111	12	US-10-083-446-18	Sequence 24, Appl
41	29	70.7	111	12	US-10-083-446-19	Sequence 9, Appl
42	29	70.7	111	12	US-10-072-571-9	Sequence 10, Appl
43	29	70.7	111	12	US-10-072-571-10	Sequence 11, Appl
44	29	70.7	111	12	US-10-072-571-11	Sequence 12, Appl
45	29	70.7	111	12	US-10-072-571-12	Sequence 12, Appl

#### ALIGNMENTS

RESULT 1  
US-09-788-308D-1  
Sequence 1, Appl  
Publication NO. US20030040468A1  
GENERAL INFORMATION:  
APPLICANT: No. US20030040468A1  
TITLE OF INVENTION: Polypeptide Pulmonary Surfactants  
FILE REFERENCE: 6374  
CURRENT APPLICATION NUMBER: US/09/788,308D  
CURRENT FILING DATE: 2001-02-16  
PRIOR APPLICATION NUMBER: US 60/182,847  
PRIOR FILING DATE: 2000-02-16  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (1)-(2)  
OTHER INFORMATION: Either Phe, Cys with an attached palmitoyl residue, or N-substituted  
OTHER INFORMATION: ted peptoid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (9)-(9)  
OTHER INFORMATION: One or more N-substituted glycine residues, such as substituted  
OTHER INFORMATION: luting but not limited to a proteinogenic amino acid side chain  
US-09-788-308D-1

Query Match 95.1% Score 39; DB 11; Length 9;  
Best Local Similarity 100.0% Pred. No. 6.1e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 PVHLKRG 9  
DB 3 PVHLKRG 9

```
RESULT 2
US-09-788-308D-4
; Sequence 4, Application US/09788308D
; Publication No. US20030040468A1
; GENERAL INFORMATION:
; APPLICANT: No. US20030040468A1Western University
; TITLE OF INVENTION: Polypeptoid Pulmonary Surfactants
; FILE REFERENCE: 6374
; CURRENT APPLICATION NUMBER: US/09/788,308D
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: US 60/182,847
; PRIOR FILING DATE: 2000-02-16
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(24)
; OTHER INFORMATION: Fifteen N-substituted glycine residues, each such residue 2-methyl
; OTHER INFORMATION: Ipropyl substituted.
US-09-788-308D-4

Query Match          95.1%; Score 39; DB 11; Length 24;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 PVHLKRG 9
Db 3 PVHLKRG 9

RESULT 3
US-09-788-308D-3
; Sequence 3, Application US/09788308D
; Publication No. US20030040468A1
; GENERAL INFORMATION:
; APPLICANT: No. US20030040468A1Western University
; TITLE OF INVENTION: Polypeptoid Pulmonary Surfactants
; FILE REFERENCE: 6374
; CURRENT APPLICATION NUMBER: US/09/788,308D
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: US 60/182,847
; PRIOR FILING DATE: 2000-02-16
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-788-308D-3

Query Match          80.5%; Score 33; DB 11; Length 35;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 PVHLKRG 8
Db 7 PVHLKRG 12

RESULT 4
US-09-864-761-41348
; Sequence 41348, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn. Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
```

```
APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 41348
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006038.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 6.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.5
; OTHER INFORMATION: SWISSPROT HIT: O14548, EVALUE 2.00e-19
; OTHER INFORMATION: EST_HUMAN HIT: AW206727.1, EVALUE 2.00e-18
US-09-864-761-41348

Query Match          80.5%; Score 33; DB 9; Length 44;
Best Local Similarity 85.7%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 PVHLKRG 9
Db 4 PVHLKRG 10

RESULT 5
US-09-984-245-278
```

Sequence 278, Application US/09984245  
Patent No. US20020165374A1  
GENERAL INFORMATION:  
APPLICANT: Young et al.  
TITLE OF INVENTION: 87 Human Secreted Proteins  
FILE REFERENCE: P2004P1  
CURRENT APPLICATION NUMBER: US/09/984,245  
CURRENT FILING DATE: 2001-10-29  
PRIOR APPLICATION NUMBER: 09/154,707  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: PCT/US98/05311  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: US 60/041,277  
PRIOR FILING DATE: 1997-03-21  
PRIOR APPLICATION NUMBER: US 60/042,344  
PRIOR FILING DATE: 1997-03-21  
PRIOR APPLICATION NUMBER: US 60/041,276  
PRIOR FILING DATE: 1997-03-21  
PRIOR APPLICATION NUMBER: US 60/041,281  
PRIOR FILING DATE: 1997-03-21  
PRIOR APPLICATION NUMBER: US 60/048,094  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,350  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,188  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,135  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/050,937  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,187  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,099  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,352  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,186  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,069  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,095  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,131  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,096  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,355  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,160  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,351  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,154  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/054,804  
PRIOR FILING DATE: 1997-08-05  
PRIOR APPLICATION NUMBER: US 60/056,370  
PRIOR FILING DATE: 1997-08-19  
PRIOR APPLICATION NUMBER: US 60/060,862  
PRIOR FILING DATE: 1997-10-02  
NUMBER OF SEQ ID NOS: 343  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 278  
LENGTH: 92  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (4)  
OTHER INFORMATION: Xaa equals any of the L-amino acids commonly found in naturally c  
US-09-984-245-278

Query Match 80.5%; Score 33; DB 10; Length 92;  
Best Local Similarity 85.7%; Pred. No. 28;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 3 PVLKRG 9  
Db 52 PVLKRG 58  
RESULT 6  
US-09-966-262-278  
Sequence 278, Application US/09966262  
Publication No. US20030050461A1  
GENERAL INFORMATION:  
APPLICANT: Young et al.  
TITLE OF INVENTION: 87 Human Secreted Proteins  
FILE REFERENCE: P2004P1  
CURRENT APPLICATION NUMBER: US/09/966,262  
CURRENT FILING DATE: 2001-10-01  
PRIOR APPLICATION NUMBER: US 09/154,707  
PRIOR FILING DATE: 1998-09-17  
PRIOR APPLICATION NUMBER: PCT/US98/05311  
PRIOR FILING DATE: 1998-03-19  
PRIOR APPLICATION NUMBER: US 60/041,277  
PRIOR FILING DATE: 1997-03-21  
PRIOR APPLICATION NUMBER: US 60/042,344  
PRIOR FILING DATE: 1997-03-21  
PRIOR APPLICATION NUMBER: US 60/041,276  
PRIOR FILING DATE: 1997-03-21  
PRIOR APPLICATION NUMBER: US 60/041,281  
PRIOR FILING DATE: 1997-03-21  
PRIOR APPLICATION NUMBER: US 60/048,094  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,350  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,188  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,135  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/050,937  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,187  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,099  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,352  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,186  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,069  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,154  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,160  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,351  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/048,154  
PRIOR FILING DATE: 1997-05-30  
PRIOR APPLICATION NUMBER: US 60/054,804  
PRIOR FILING DATE: 1997-08-05  
PRIOR APPLICATION NUMBER: US 60/056,370  
PRIOR FILING DATE: 1997-08-19  
PRIOR APPLICATION NUMBER: US 60/060,862  
PRIOR FILING DATE: 1997-10-02  
NUMBER OF SEQ ID NOS: 343  
SOFTWARE: PatentIn Ver. 2.0

```
/ SEQ ID NO 278
/ LENGTH: 92
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: MISC_FEATURE
/ LOCATION: (4)
/ OTHER INFORMATION: Xaa equals any of the L-amino acids commonly found in naturally
/ OTHER INFORMATION: proteins
US-09-966-262-278
```

```
Query Match      80.5%; Score 33; DB 11; Length 92;
Best Local Similarity 85.7%; Pred. No. 28;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      3 PVLKRG 9
        ||:||||
Db       52 PVLKRG 58
```

```
RESULT 7
US-09-983-966-278
/ Sequence 278, Application US/09983966
/ Publication No. US20030060619A1
/ GENERAL INFORMATION:
/ APPLICANT: Young et al.
/ TITLE OF INVENTION: 87 Human Secreted Proteins
/ FILE REFERENCE: P2004P1
/ CURRENT APPLICATION NUMBER: US/09/983,966
/ PRIOR FILING DATE: 2001-10-29
/ PRIOR APPLICATION NUMBER: 09/154,707
/ PRIOR FILING DATE: 1998-09-17
/ PRIOR APPLICATION NUMBER: PCT/US98/05311
/ PRIOR FILING DATE: 1998-03-19
/ PRIOR APPLICATION NUMBER: US 60/041,277
/ PRIOR FILING DATE: 1997-03-21
/ PRIOR APPLICATION NUMBER: US 60/042,344
/ PRIOR FILING DATE: 1997-03-21
/ PRIOR APPLICATION NUMBER: US 60/041,276
/ PRIOR FILING DATE: 1997-03-21
/ PRIOR APPLICATION NUMBER: US 60/041,281
/ PRIOR FILING DATE: 1997-03-21
/ PRIOR APPLICATION NUMBER: US 60/048,094
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/048,350
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/048,188
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/048,135
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/050,937
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/048,187
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/048,099
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/048,352
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/048,186
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/048,069
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/048,095
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/048,131
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/048,096
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/048,355
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/048,160
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/048,351
```

```
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/048,154
/ PRIOR FILING DATE: 1997-05-30
/ PRIOR APPLICATION NUMBER: US 60/054,804
/ PRIOR FILING DATE: 1997-08-05
/ PRIOR APPLICATION NUMBER: US 60/056,370
/ PRIOR FILING DATE: 1997-08-19
/ PRIOR APPLICATION NUMBER: US 60/060,862
/ PRIOR FILING DATE: 1997-10-02
/ NUMBER OF SEQ ID NOS: 343
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 278
/ LENGTH: 92
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: MISC_FEATURE
/ LOCATION: (4)
/ OTHER INFORMATION: Xaa equals any of the L-amino acids commonly found in naturally
/ OTHER INFORMATION: proteins
US-09-983-966-278
```

```
Query Match      80.5%; Score 33; DB 11; Length 92;
Best Local Similarity 85.7%; Pred. No. 28;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      3 PVLKRG 9
        ||:||||
Db       52 PVLKRG 58
```

```
RESULT 8
US-10-143-090-278
/ Sequence 278, Application US/10143090
/ Publication No. US20030069406A1
/ GENERAL INFORMATION:
/ APPLICANT: Young et al.
/ TITLE OF INVENTION: 87 Human Secreted Proteins
/ FILE REFERENCE: P2004P1
/ CURRENT APPLICATION NUMBER: US/10/143,090
/ PRIOR FILING DATE: 2002-05-13
/ PRIOR APPLICATION NUMBER: 09/154,707
/ PRIOR FILING DATE: 1998-09-17
/ NUMBER OF SEQ ID NOS: 343
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 278
/ LENGTH: 92
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: MISC_FEATURE
/ LOCATION: (4)
/ OTHER INFORMATION: Xaa equals any of the L-amino acids commonly found in naturally
US-10-143-090-278
```

```
Query Match      80.5%; Score 33; DB 15; Length 92;
Best Local Similarity 85.7%; Pred. No. 28;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      3 PVLKRG 9
        ||:||||
Db       52 PVLKRG 58
```

```
RESULT 9
US-10-074-247-2
/ Sequence 2, Application US/10074247
/ Publication No. US20020197646A1
/ GENERAL INFORMATION:
/ APPLICANT: No. US20020197646A1ee, Lawrence M.
/ APPLICANT: Whitsett, Jeffrey A.
/ APPLICANT: Cole, F. Sessions
/ APPLICANT: Hamvas, Aaron
```

TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated with Interstitial Lung  
FILE REFERENCE: 001107.00229  
CURRENT APPLICATION NUMBER: US/10/074,247  
CURRENT FILING DATE: 2002-02-14  
PRIOR APPLICATION NUMBER: US 60/268,650  
PRIOR FILING DATE: 2001-02-14  
PRIOR APPLICATION NUMBER: US 60/268,991  
PRIOR FILING DATE: 2001-02-15  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 197  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-074-247-2

Query Match 80.5%; Score 33; DB 14; Length 197;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 PVHLKR 8  
Db 30 PVHLKR 35

RESULT 10  
US-09-925-302-482  
Sequence 482, Application US/09925302  
Patent No. US20020044941A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA104  
CURRENT APPLICATION NUMBER: US/09/925,302  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05918  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 896  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 482  
LENGTH: 216  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (5)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (8)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-302-482

Query Match 80.5%; Score 33; DB 9; Length 216;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 PVHLKR 8  
Db 55 PVHLKR 60

RESULT 11  
US-09-925-301-1016  
Sequence 1016, Application US/09925301  
Patent No. US20020052308A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA106  
CURRENT APPLICATION NUMBER: US/09/925,301  
CURRENT FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05882  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1694  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 1016  
LENGTH: 257  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-925-301-1016

Query Match 80.5%; Score 33; DB 9; Length 257;  
Best Local Similarity 85.7%; Pred. No. 74;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 PVHLKRG 9  
Db 115 PVHLTRG 121

RESULT 12  
US-10-023-896-59  
Sequence 59, Application US/10023896  
Publication No. US2003002776A1  
GENERAL INFORMATION:  
APPLICANT: Victor Roschke  
TITLE OF INVENTION: 29 Human Cancer Associated Proteins  
FILE REFERENCE: PA004P1  
CURRENT APPLICATION NUMBER: US/10/023,896  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: unassigned  
PRIOR FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: PCT/US00/23794  
PRIOR FILING DATE: 2000-08-30  
PRIOR APPLICATION NUMBER: 60/152,296  
PRIOR FILING DATE: 1999-09-03  
PRIOR APPLICATION NUMBER: 60/158,003  
PRIOR FILING DATE: 1999-10-06  
NUMBER OF SEQ ID NOS: 138  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 59  
LENGTH: 257  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-023-896-59

Query Match 80.5%; Score 33; DB 15; Length 257;  
Best Local Similarity 85.7%; Pred. No. 74;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 PVHLKRG 9  
Db 115 PVHLTRG 121

RESULT 13  
US-10-023-896-87  
Sequence 87, Application US/10023896  
Publication No. US2003002776A1  
GENERAL INFORMATION:  
APPLICANT: Victor Roschke  
TITLE OF INVENTION: 29 Human Cancer Associated Proteins  
FILE REFERENCE: PA004P1  
CURRENT APPLICATION NUMBER: US/10/023,896  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: unassigned  
PRIOR FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: PCT/US00/23794  
PRIOR FILING DATE: 2000-08-30  
PRIOR APPLICATION NUMBER: 60/152,296  
PRIOR FILING DATE: 1999-09-03  
PRIOR APPLICATION NUMBER: 60/158,003

PRIOR FILING DATE: 1999-10-06  
NUMBER OF SEQ ID NOS: 138  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 87  
LENGTH: 257  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-023-896-87

Query Match 80.5%; Score 33; DB 15; Length 257;  
Best Local Similarity 85.7%; Pred. No. 74;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 PVHLKRG 9  
|||  
Db 115 PVHLTRG 121

RESULT 14  
US-10-116-519-8  
Sequence 8, Application US/10116519  
Publication No. US20030114373A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDE ENCODING A NOVEL CYSTEINE PROTEASE OF THE CALPAIN  
FILE REFERENCE: D0124 NP  
CURRENT APPLICATION NUMBER: US/10/116,519  
PRIOR FILING DATE: 2002-04-03  
PRIOR APPLICATION NUMBER: US 60/281,253  
PRIOR FILING DATE: 2001-04-03  
PRIOR APPLICATION NUMBER: US 60/288,768  
PRIOR FILING DATE: 2001-05-04  
PRIOR APPLICATION NUMBER: US 60/296,180  
PRIOR FILING DATE: 2001-06-06  
PRIOR APPLICATION NUMBER: US 60/300,620  
PRIOR FILING DATE: 2001-06-25  
NUMBER OF SEQ ID NOS: 145  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO: 8  
LENGTH: 713  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-116-519-8

Query Match 80.5%; Score 33; DB 15; Length 713;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKR 8  
|||  
Db 456 PVHLKR 461

RESULT 15  
US-10-274-878-2  
Sequence 2, Application US/10274878  
Publication No. US20030049792A1  
GENERAL INFORMATION:  
APPLICANT: RUSCH, Douglas et al  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
FILE REFERENCE: CLO01099-CIP-DIV  
CURRENT APPLICATION NUMBER: US/10/274,878  
PRIOR FILING DATE: 2002-10-22  
PRIOR APPLICATION NUMBER: 09/849,334  
PRIOR FILING DATE: 2001-05-07  
PRIOR APPLICATION NUMBER: 09/773,371  
PRIOR FILING DATE: 2001-02-01  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 2

LENGTH: 723  
TYPE: PRT  
ORGANISM: Human  
US-10-274-878-2

Query Match 80.5%; Score 33; DB 15; Length 723;  
Best Local Similarity 85.7%; Pred. No. 2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 PVHLKRG 9  
|||  
Db 566 PVHLTRG 572

Search completed: December 3, 2003, 15:56:07  
Job time : 5.44898 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 15:45:30 / Search time 2.63265 Seconds  
(without alignments)  
328.763 Million cell updates/sec

Title: US-09-788-308d-1

Perfect score: 41  
Sequence: 1 XXPVHLKRG 9

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

1: PIR\_76:\*  
2: PIR:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	87.8	373	2 AH0693	conserved hypotet
2	34	82.9	367	2 S72924	hypothetical prote
3	34	82.9	370	2 G87212	conserved hypotet
4	34	82.9	492	1 B69114	fumarate reductase
5	34	82.9	847	2 A56039	GTPase-activating
6	34	82.9	850	2 JC5047	ras GTPase-activat
7	33	80.5	189	1 LNRBC1	pulmonary surfacta
8	33	80.5	191	2 G02964	surfactant protein
9	33	80.5	193	1 A36534	pulmonary surfacta
10	33	80.5	194	1 LNRPC	pulmonary surfacta
11	33	80.5	197	1 LNRUC	pulmonary surfacta
12	33	80.5	294	1 S23561	HAL1 protein - yea
13	33	80.5	494	2 T14729	betaine-aldehyde d
14	33	80.5	505	2 T03394	probable betaine-a
15	32	78.0	80	1 OSB07A	cytochrome-c oxida
16	32	78.0	83	2 I48286	cytochrome-c oxida
17	32	78.0	83	2 A49355	cytochrome-c oxida
18	32	78.0	83	2 S13099	cytochrome-c oxida
19	32	78.0	507	2 G01614	zinc finger protei
20	32	78.0	511	2 T07787	pyruvate kinase (E
21	32	78.0	628	2 A81999	glucose inhibited
22	32	78.0	631	2 F81227	glucose inhibited
23	32	78.0	913	2 S61580	probable membrane
24	32	78.0	1042	2 H70203	isooleucine-tRNA 11
25	32	78.0	1513	2 A54895	mucin 2, intestina
26	32	78.0	3020	2 A43932	mucin 2 precursor
27	32	78.0	3343	2 S44887	ZK112.7 protein -
28	31	75.6	101	2 C72467	hypothetical prote
29	31	75.6	139	2 T16449	pathogenicity fact

30	31	75.6	185	2 D82772	phage-related tail
31	31	75.6	247	2 T31140	hypothetical prote
32	31	75.6	271	2 T18056	ATPase homolog A55
33	31	75.6	346	2 S49963	hypothetical prote
34	31	75.6	472	2 B26580	chromosome replica
35	31	75.6	489	1 D64311	fumarate reductase
36	31	75.6	538	2 B83625	probable gamma-glu
37	31	75.6	574	2 S50800	probable membrane
38	31	75.6	736	1 M4XR3D	structural protein
39	31	75.6	736	2 S23554	Mu2 protein - reov
40	31	75.6	750	2 T04010	hypothetical prote
41	31	75.6	1324	2 T18265	endo-1,3(4)-beta-g
42	30	73.2	114	2 A83300	hypothetical prote
43	30	73.2	170	2 D87707	hypothetical prote
44	30	73.2	183	2 T18193	hypothetical prote
45	30	73.2	183	2 T17492	hypothetical prote

#### ALIGNMENTS

RESULT 1  
AH0693  
conserved hypothetical protein STY1676 [imported] - Salmonella enterica subsp. enterica  
C/Species: Salmonella enterica subsp. enterica serovar Typhi  
A/Note: this species has also been called Salmonella typhi  
C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C/Accession: AH0693  
R/Parikh, V.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero A/Reference number: AB0502; MUID:21534947; PMID:11677608  
A/Accession: AH0693  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-373 <PAR>  
A/Cross-references: GB:AL513382; PIDN:CAD01921.1; PID:gl6502763; GSPDB:GN00176  
C/Genetics:  
A/Gene: STY1676  
C/Superfamily: hypothetical protein HI0753

Query Match 87.8%; Score 36; DB 2; Length 373;  
Best Local Similarity 85.7%; Pred: No. 9.5;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKRG 9  
DB 41 PVHLKRG 47

RESULT 2  
S72924  
hypothetical protein B2168 Cl 175 - Mycobacterium leprae  
C/Species: Mycobacterium leprae  
C/Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 23-Mar-2001  
C/Accession: S72924  
R/Smith, D.R.; Robison, K.  
submitted to the EMBL Data Library, November 1993  
A/Description: Mycobacterium leprae cosmid B2168.  
A/Reference number: S72586  
A/Accession: S72924  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-367 <SMI>  
A/Cross-references: EMBL:U00018; NID:g467037; PIDN:AAA17260.1; PID:g467076  
C/Genetics:  
A/Start codon: GTG  
C/Superfamily: Streptomyces coelicolor hypothetical protein SCE68.23c

Query Match 82.9%; Score 34; DB 2; Length 367;



Best Local Similarity 71.4%; Pred. No. 24;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKRG 9  
|:|:|:|  
Db 19 PVHLKRG 25

## RESULT 3

Conserved hypothetical protein ML2427 [imported] - Mycobacterium leprae  
C/Species: Mycobacterium leprae  
C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C/Accession: G87212  
R/Col: S.T.; Eiglsmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho  
R.; Davies, R.M.; Devlin, K.; Duthey, S.; Felwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A/Author: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq  
A/Title: Massive gene decay in the leprosy bacillus.  
A/Reference number: A86909; MUID:21128732; PMID:11234002  
A/Accession: G87212  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-370 <STO>  
A/Cross-references: GB:AL450380; NID:g13094001; PIDN:CAC31943.1; GSPDB:GN00147  
C/Genetics:  
A/Supernfamily: Streptomyces coelicolor hypothetical protein SCE68.23c

Query Match 82.9%; Score 34; DB 2; Length 370;  
Best Local Similarity 71.4%; Pred. No. 24;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKRG 9  
|:|:|:|  
Db 22 PVHLKRG 28

## RESULT 4

E69114  
fumarate reductase (EC 1.3.99.1) - Methanobacterium thermoautotrophicum (strain Delta H)  
C/Species: Methanobacterium thermoautotrophicum  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 28-Jul-2000  
C/Accession: E69114  
R/Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
; Qiu, D.; Spadefora, R.; Vicaire, R.; Wang, Y.; Mierzdowski, J.; Gibson, R.; Jiwani, N.  
K.; S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func  
A/Reference number: A69000; MUID:98037514; PMID:9371463  
A/Accession: E69114  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-492 <MTA>  
A/Cross-references: GB:AE000937; GB:AE000666; NID:g2622974; PIDN:AAB6316.1; PID:g262298  
A/Experimental source: strain Delta H  
C/Genetics:  
A/Gene: MTH1850  
C/Supernfamily: Methanococcus fumarate reductase; ferredoxin [2Fe-2S] homology  
C/Keywords: 2Fe-2S; metalloprotein; oxidoreductase  
F/36-78/Domain: ferredoxin [2Fe-2S] homology <FER1>  
F/57,62,65,77/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 82.9%; Score 34; DB 1; Length 492;  
Best Local Similarity 85.7%; Pred. No. 32;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 PVHLKRG 9  
|:|:|:|  
Db 386 PVHLKRG 392

Best Local Similarity 82.9%; Score 34; DB 2; Length 847;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 PVHLKRG 9  
|:|:|:|  
Db 602 PVHLKRG 608

## RESULT 5

A56039  
GTPase-activating protein Gap1(m) - rat.  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 01-Dec-1995 #sequence\_revision 01-Dec-1995 #text\_change 05-Nov-1999  
C/Accession: A56039  
R/Makawa, M.; Li, S.; Iwamatsu, A.; Morishita, T.; Yokota, K.; Imai, Y.; Kohsaka, S.; I  
Mol. Cell. Biol. 14, 6879-6885, 1994  
A/Title: A novel mammalian Ras GTPase-activating protein which has phospholipid-binding  
A/Reference number: A56039; MUID:95021216; PMID:7935405  
A/Accession: A56039  
A/Status: preliminary; nucleic acid sequence not shown  
A/Molecule type: mRNA  
A/Residues: 1-847 <MAE>  
A/Cross-references: GB:D0734; NID:g553374; PIDN:BAA06398.1; PID:d1006967; PID:g553375  
A/Experimental source: brain  
C/Supernfamily: ras-specific GAP catalytic domain homology; pleckstrin repeat homology  
F/355-567/Domain: ras-specific GAP catalytic domain homology <GAP>  
F/602-702/Domain: pleckstrin repeat homology <PLK>

Query Match 82.9%; Score 34; DB 2; Length 847;  
Best Local Similarity 85.7%; Pred. No. 54;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 PVHLKRG 9  
|:|:|:|  
Db 602 PVHLKRG 608

## RESULT 6

JC5047  
ras GTPase-activating protein - human.  
C/Species: Homo sapiens (man)  
C/Date: 31-Jan-1997 #sequence\_revision 31-Jan-1997 #text\_change 05-Nov-1999  
C/Accession: JC5047  
R/Kobayashi, M.; Masui, T.; Kusuda, J.; Kameoka, Y.; Hashimoto, K.; Iwashita, S.  
Gene 175, 173-177, 1996  
A/Title: Human rasGTPase-activating protein (human counterpart of Gap1m): Sequence of t  
A/Reference number: JC5047; MUID:97074668; PMID:8917095  
A/Accession: JC5047  
A/Molecule type: mRNA  
A/Residues: 1-850 <KOB>  
A/Cross-references: DBJ:D78155; NID:g1060908; PIDN:BAI1230.1; PID:d1011892; PID:g1060  
C/Comment: This protein plays a role in the regulation of cell growth and differentiat  
C/Genetics:  
A/Gene: GAP1m  
A/Map position: 3q24-26  
C/Supernfamily: pleckstrin repeat homology; ras-specific GAP catalytic domain homology  
F/356-568/Domain: ras-specific GAP catalytic domain homology <GAP>  
F/603-704/Domain: pleckstrin repeat homology <PLK>

Query Match 82.9%; Score 34; DB 2; Length 850;  
Best Local Similarity 85.7%; Pred. No. 54;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 PVHLKRG 9  
|:|:|:|  
Db 603 PVHLKRG 609

## RESULT 7

LNRBC1  
pulmonary surfactant protein C precursor - rabbit  
N/Alternate names: surfactant-associated protein SP-C  
C/Species: Oryctolagus cuniculus (domestic rabbit)  
C/Date: 01-Sep-1995 #sequence\_revision 25-Apr-1997 #text\_change 18-Feb-2000  
C/Accession: A56766; S14815; A56860; S19946; A56655  
R/Boggarlam, V.; Margana, R.K.  
Am. J. Physiol. 263, L634-L644, 1992  
A/Title: Rabbit surfactant protein C: cDNA cloning and regulation of alternatively spli  
A/Reference number: A56766; MUID:9318799; PMID:1335697  
A/Accession: A56766  
A/Molecule type: mRNA

A/Residues: 1-189 <BOG>  
A/Cross-references: GB:S51993; NID:G262767; PIDN:AA24761.1; PID:G262768; GB:S51997; NID  
A/Experimental source: lung  
A/Note: sequence extracted from NCBI backbone (NCBIN:121728, NCBI:P:121729)  
A/Note: two types of mRNA were found differing in their 3' untranslated regions because  
R:Johnson, J.; Peterson, P.; Loewenadler, B.; Robertson, B.; Joernvall, H.; Cursedt, T  
FEBS Lett. 281, 119-122, 1991  
A/Title: Canine hydrophobic surfactant polypeptide SP-C. A lipopeptide with one thioester  
A/Reference number: S14813; MUID:91200266; PMID:2015882  
A/Accession: S14815  
A/Molecule type: protein  
A/Residues: 24-37, XXXXXXXXXXXXXXXX <JOH>  
R:Connell, I.; Possmayer, F.  
Biochim. Biophys. Acta 1127, 199-207, 1992  
A/Title: cDNA sequence and alternative mRNA splicing of surfactant-associated protein C  
A/Reference number: A56860; MUID:92353123; PMID:1643107  
A/Accession: A56860  
A/Molecule type: mRNA  
A/Residues: 24-58 <CON>  
A/Cross-references: EMBL:X65078; NID:G1720; PIDN:CAA46204.1; PID:G1721  
A/Experimental source: fetal lung  
A/Note: sequence extracted from NCBI backbone (NCBIN:110198, NCBI:P:110234); the complete  
R:Connell, I.; Possmayer, F.  
submitted to the EMBL Data Library, March 1992  
A/Description: cDNA sequence and alternative splicing of surfactant-associated protein C  
A/Reference number: S19946  
A/Accession: S19946  
A/Molecule type: mRNA  
A/Residues: 3-115, 117-161, 'R', 163-189 <CO2>  
A/Cross-references: EMBL:X65078; NID:G1720; PIDN:CAA46204.1; PID:G1721  
R:Durrham, P.L.; Nanthakumar, E.J.; Snyder, J.M.  
Exp. Lung Res. 18, 775-793, 1992  
A/Title: Developmental regulation of surfactant-associated proteins in rabbit fetal lung  
A/Reference number: A56555; MUID:93105936; PMID:1468410  
A/Accession: A56555  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 24-115, 117-186, 'Y', 188-189 <DUR>  
A/Cross-references: GB:S51098; NID:G262066; PIDN:AA24576.1; PID:G262067  
A/Experimental source: fetal lung  
A/Note: sequence extracted from NCBI backbone (NCBIN:121095, NCBI:P:121096); sequence ind  
C/Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers t  
C/Comment: This protein is synthesized by alveolar type II cells.  
C/Superfamily: pulmonary surfactant protein C  
C/Keywords: gaseous exchange; lipoprotein; lung; pulmonary surfactant; thiolester bond  
F:1-23/Domain: propeptide #status predicted <PRO>  
F:24-58/Product: pulmonary surfactant protein C #status experimental <MAT>  
F:28,29/Binding site: palmitate (Cys) (covalent) #status experimental

Query Match 80.5%; Score 33; DB 1; Length 189;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKR 8  
Db 30 PVHLKR 35

RESULT 8  
G02864  
surfactant protein C - rhesus macaque  
C/Species: Macaca mulatta (rhesus macaque)  
C/Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 20-Aug-1999  
C/Accession: G02864  
R/An, G.  
submitted to the EMBL Data Library, February 1994  
A/Reference number: G12797  
A/Accession: G02864  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-191 <ANX>  
A/Cross-references: EMBL:U06694; NID:G476267; PIDN:AAA17870.1; PID:G476268

C/Superfamily: pulmonary surfactant protein C

Query Match 80.5%; Score 33; DB 2; Length 191;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKR 8  
Db 30 PVHLKR 35

RESULT 9  
A36534  
pulmonary surfactant protein C precursor - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 12-Apr-1991 #sequence\_revision 25-Apr-1997 #text\_change 18-Feb-2000  
C/Accession: A36534  
R:Glaser, S.W.; Korfhagen, T.R.; Bruno, M.D.; Dey, C.; Whiteett, J.A.  
J. Biol. Chem. 265, 21986-21991, 1990  
A/Title: Structure and expression of the pulmonary surfactant protein SP-C gene in the r  
A/Reference number: A36534; MUID:91072410; PMID:2254341  
A/Accession: A36534  
A/Molecule type: DNA  
A/Residues: 1-193 <GLA>  
A/Cross-references: GB:M38314; NID:G200559; PIDN:AAA40010.1; PID:G200560  
C/Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers t  
C/Comment: This protein is synthesized by alveolar type II cells.  
C/Superfamily: pulmonary surfactant protein C  
C/Keywords: gaseous exchange; lipoprotein; lung; pulmonary surfactant; thiolester bond  
F:1-23/Domain: propeptide #status predicted <PRO>  
F:24-58/Product: pulmonary surfactant protein C #status predicted <MAT>  
F:28,29/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 80.5%; Score 33; DB 1; Length 193;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKR 8  
Db 30 PVHLKR 35

RESULT 10  
LNRC  
pulmonary surfactant protein C precursor - rat  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 18-Feb-2000  
C/Accession: S03994  
R:fisher, J.H.; Shannon, J.M.; Hofmann, T.; Mason, R.J.  
Biochim. Biophys. Acta 995, 225-230, 1988  
A/Title: Nucleotide and deduced amino acid sequence of the hydrophobic surfactant prote  
A/Reference number: S03994; MUID:8920572; PMID:2706272  
A/Accession: S03994  
A/Molecule type: mRNA  
A/Residues: 1-194 <FIS>  
A/Cross-references: GB:X14221; NID:G57286; PIDN:CAA32440.1; PID:G57287  
C/Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers t  
C/Comment: This protein is synthesized by alveolar type II cells.  
C/Superfamily: pulmonary surfactant protein C  
C/Keywords: gaseous exchange; lipoprotein; lung; pulmonary surfactant; thiolester bond  
F:1-23/Domain: propeptide #status predicted <PRO>  
F:24-58/Product: pulmonary surfactant protein C #status predicted <MAT>  
F:28,29/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 80.5%; Score 33; DB 1; Length 194;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKR 8  
Db 30 PVHLKR 35

## RESULT 11

LNHUC

pulmonary surfactant protein C precursor, long splice form [validated] - human  
 N:Alternate names: 3.7 kDa surfactant polypeptide; pulmonary surfactant protein SP5; pul  
 N:Contains: pulmonary surfactant protein C precursor, short splice form  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1990 #sequence revision 31-Dec-1990 #text change 08-Dec-2000  
 A:Accession: A28801; EMBL:028801; F38420; A27338; S02315; S00608; A61249  
 R:Glasner, S.W.; Korfhaugen, T.R.; Perme, C.M.; Pilot-Matias, T.J.; Kister, S.E.; Whitee  
 U. Biol. Chem. 263, 10326-10331, 1988  
 A>Title: Two SP-C genes encoding human pulmonary surfactant proteolipid.  
 A:Reference number: A28801; MUID:88273133; PMID:2839484  
 A:Accession: A28801  
 A:Molecule type: DNA  
 A:Residues: 1-137, 'T', 139-197 <GLA1>  
 A:Cross-references: GB:J03890; NID:G190089; PIDN:AA32022.1; PID:G387029  
 A:Experimental source: long splice form  
 A>Note: the codon given for residue 138 (ATT) is inconsistent with the authors' translat  
 A:Accession: B28801  
 A:Molecule type: DNA  
 A:Residues: 1-137, 'T', 139-145, 152-197 <GLA2>  
 A:Cross-references: GB:J03890; NID:G190089; PIDN:AA32023.1; PID:G387030  
 A:Experimental source: short splice form  
 A>Note: the codon given for residue 138 (ATT) is inconsistent with the authors' translat  
 A>Note: identical proteins are encoded by two genes; transcripts from either gene can be  
 R:Hatzis, D.; Delter, G.; demello, D.B.; Floros, J.  
 Exp. Lung Res. 20, 57-72, 1994  
 A>Title: Human surfactant protein-C: genetic homogeneity and expression in RDS; comparis  
 A:Reference number: 138420; MUID:94237133; PMID:8181452  
 A:Accession: 138420  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-14, 'PCO', 15-44, 'S', 46-64, 'FPQ', 68-137, 'T', 139-185, 'S', 187-197 <HAT>  
 A:Cross-references: EMBL:U02948; NID:G498319; PIDN:AA860332.1; PID:G514214  
 A>Note: there are probably errors in the presentation of the CDS splice boundaries in ch  
 R:Glasner, S.W.; Korfhaugen, T.R.; Weaver, T.E.; Clark, J.C.; Pilot-Matias, T.; Meuth, J.  
 J. Biol. Chem. 263, 9-12, 1988  
 A>Title: CDNA, deduced polypeptide structure and chromosomal assignment of human pulmona  
 A:Reference number: A27338; MUID:88087156; PMID:3335510  
 A:Accession: A27338  
 A:Molecule type: mRNA  
 A:Residues: 1-197 <GLA3>  
 A:Cross-references: GB:J03517; NID:G338412; PIDN:AAA6634.1; PID:G338413  
 A:Note: part of this sequence, including the amino end of the mature protein, was deter  
 R:Warr, R.G.; Hawgood, S.; Buckley, D.T.; Crisp, T.M.; Schilling, J.; Benson, B.J.; Ball  
 Proc. Natl. Acad. Sci. U.S.A. 84, 7915-7919, 1987  
 A>Title: Low molecular weight human pulmonary surfactant protein (SP5): isolation, chara  
 A:Reference number: S02315; MUID:88068508; PMID:3479771  
 A:Accession: S02315  
 A:Molecule type: mRNA  
 A:Residues: 1-197 <MAR>  
 A:Cross-references: GB:J03553; NID:G338306; PIDN:AAA6631.1; PID:G338307  
 A:Note: part of this sequence, including the amino end of the mature protein, was confir  
 R:Johnanson, J.; Joernvall, H.; Eklund, A.; Christensen, N.; Robertsson, B.; Cursstedt, T.  
 FEBS Lett. 232, 61-64, 1988  
 A>Title: Hydrophobic 3.7 kDa surfactant polypeptide: structural characterization of the  
 A:Reference number: S00608; MUID:88211876; PMID:3366248  
 A:Accession: S00608  
 A:Molecule type: protein  
 A:Residues: 24-58 <JOH>  
 A:Note: 25-Arg was also found  
 A:Note: peptides beginning at residues 24, 25, and 26 were detected  
 R:Stults, J.T.; Griffin, P.R.; Leskar, D.D.; Naidu, A.; Moffat, B.; Benson, B.J.  
 Am. J. Physiol. 261, L118-L125, 1991  
 A>Title: Lung surfactant protein SP-C from human, bovine, and canine sources contains pa  
 A:Reference number: A61249; MUID:91336436; PMID:1872406  
 A:Accession: A61249  
 A:Molecule type: protein  
 A:Residues: 24-58 <STU>  
 A:Note: identification of palmitoyl cysteines

C:Comment: Pulmonary surfactant is a complex of phospholipide and proteins that lowers

C:Comment: This protein is synthesized by alveolar type II cells.

C:Comment: The precursor of the hydrophobic proteolipid protein C lacks a typical signa

C:Genetics:

A:Gene: GDB:SFTPC; SFTP2; SP-C

A:Cross-references: GDB:120373; OMIM:178620

A:Map position: 8p21-8p21

A:Introns: 14/3; 67/3; 108/3; 145/3

A:Note: the first intron occurs before the initiator codon

C:Superfamily: pulmonary surfactant protein C

C:Keywords: alternative splicing; gaseous exchange; lipoprotein; lung; pulmonary surfac

F:1-197/Product: pulmonary surfactant protein C precursor, short splice form #status pr

F:1-145,152-197/Product: pulmonary surfactant protein C precursor, short splice form #s

F:1-23/Domain: propeptide #status predicted <PRO>

F:24-58/Product: pulmonary surfactant protein C #status experimental <MAT>

F:28,29/Binding site: palmitate (Cys) (covalent) #status experimental

## Query Match

Best Local Similarity 80.5%; Score 33; DB 1; Length 197;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKRG 8

DB 30 PVHLKRG 35

## RESULT 12

HAL1 protein - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein LP25c; protein YP9723\_05c; protein YPR005c

C:Species: Saccharomyces cerevisiae

C:Date: 12-Mar-1993 #sequence revision 12-Mar-1993 #text change 04-Mar-2000

C:Accession: S23561; S52818; S59750

R:Gaxiola, R.; de Larrinoa, I.F.; Villalba, J.M.; Serrano, R.

EMBO J. 11, 3157-3164, 1992

A>Title: A novel and conserved salt-induced protein is an important determinant of salt

A:Reference number: S23561; MUID:92371421; PMID:1505513

A:Accession: S23561

A:Molecule type: DNA

A:Residues: 1-294 <GAX>

A:Cross-references: EMBL:X67559; NID:G3760; PIDN:CAA47858.1; PID:G3761

A:Experimental source: strain RS-16

R:Pearson, D.; Bowman, S.

submitted to the EMBL Data Library, April 1995

A:Reference number: S52814

A:Accession: S52818

A:Molecule type: DNA

A:Residues: 1-294 <PEA>

A:Cross-references: EMBL:Z48951; NID:G76299; PIDN:CAA88783.1; PID:G763004; MIPS:YPR005

A:Experimental source: strain AB972

R:Wang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Hall, J.; Storms, R.K.; V

submitted to the EMBL Data Library, July 1995

A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI right arm.

A:Reference number: S59746

A:Accession: S59750

A:Molecule type: DNA

A:Residues: 1-294 <MAN>

A:Cross-references: EMBL:U31900; NID:G1276597; PIDN:AAA97584.1; PID:G939739; MIPS:YPR00

C:Genetics:

A:Gene: SGD:HAL1

A:Cross-references: SGD:S0006209; MIPS:YPR005c

A:Map position: 16R

C:Superfamily: Saccharomyces cerevisiae HAL1 protein

## Query Match

Best Local Similarity 80.5%; Score 33; DB 2; Length 294;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 PVHLKRG 9

DB 112 PVHLKRG 118

RESULT 13  
T14729  
C:Accession: A41852; S18187; A25879; S13002; A41034  
C:Species: Sorghum bicolor (sorghum)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: T14729  
R:Wood, A.U.; Saneoka, H.; Joly, R.J.; Rhodes, D.; Goldbrough, P.B.  
A:Title: Betaine aldehyde dehydrogenase in Sorghum bicolor: molecular cloning and expres  
A:Reference number: Z18171; MUID:97088719; PMID:8934627  
A:Accession: T14729  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-494 <W00>  
A:Cross-references: EMBL:U12196; NID:9520545; PIDN:AAC49268.1; PID:9520546  
A:Experimental source: cultivar P954035; leaf, stem  
C:Function:  
A:Description: catalyzes the oxidation of betaine aldehyde to betaine using NAD+ and wat  
A:Pathway: betaine biosynthesis  
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology  
C:Keywords: NAD; oxidoreductase  
F:42-308/Domain: aldehyde dehydrogenase homology <ALD>

Query Match 80.5%; Score 33; DB 2; Length 494;  
Best Local Similarity 85.7%; Pred. No. 52;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

3 PVHLKRG 9  
| | | | |  
Db 360 PQHLKRG 366

RESULT 14  
T03394  
C:Species: Oryza sativa (rice)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000  
C:Accession: T03394  
R:Nakamura, T.; Yokota, S.; Muramoto, Y.; Tautsui, K.; Oguri, Y.; Fukui, K.; Takabe, T.  
A:Title: Expression of a betaine aldehyde dehydrogenase gene in rice, a glycine betaine  
A:Reference number: Z14925; MUID:97336302; PMID:9193078  
A:Accession: T03394  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-505 <NNA>  
A:Cross-references: EMBL:AB001348; NID:92244603; PIDN:BA421098.1; PID:92244604  
A:Experimental source: cv. Nipponbare  
C:Genetics:  
A:Insertions: 38/1; 85/3; 112/3; 162/2; 193/2; 234/3; 256/3; 279/1; 304/3; 342/3; 368/3; 41  
C:Function:  
A:Description: catalyzes the oxidation of betaine aldehyde to betaine using NAD+ and wat  
A:Pathway: betaine biosynthesis  
A>Note: betaine is a protective osmolyte induced to accumulate under saline or dry condi  
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology  
C:Keywords: NAD; oxidoreductase; stress-induced protein  
F:48-316/Domain: aldehyde dehydrogenase homology <ALD>

Query Match 80.5%; Score 33; DB 2; Length 505;  
Best Local Similarity 85.7%; Pred. No. 53;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

3 PVHLKRG 9  
| | | | |  
Db 368 PQHLKRG 374

RESULT 15  
OSB07A  
C:cytochrome-c oxidase (EC 1.9.3.1) chain VIIa precursor, cardiac [validated] - bovine  
N:Alternate names: cytochrome-c oxidase chain VIIa, isoform H; cytochrome-c oxidase cha  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 30-Sep-1993 #sequence\_revision 07-Jun-1996 #text\_change 15-Sep-2000

C:Accession: A41852; S18187; A25879; S13002; A41034  
R:Seelan, R.S.; Grossman, L.I.  
A:Title: Structure and organization of the heart isoform gene for bovine cytochrome c o  
A:Reference number: A41852; MUID:92256407; PMID:1316159  
A:Accession: A41852  
A:Molecule type: DNA  
A:Residues: 1-80 <SEE>  
A:Cross-references: GB:M63299; NID:9162894; PIDN:AAA30464.1; PID:9162895  
R:Seelan, R.S.; Grossman, L.I.  
J. Biol. Chem. 266, 19752-19757, 1991  
A:Title: Cytochrome c oxidase subunit VIIa isoforms. Characterization and expression of  
A:Reference number: S18187; MUID:92011781; PMID:1717471  
A:Accession: S18187  
A:Molecule type: mRNA  
A:Residues: 1-80 <SE2>  
A:Cross-references: GB:X56739; NID:9271; PIDN:CAA40063.1; PID:9272  
R:Meinecke, L.; Buse, G.  
Biol. Chem. Hoppe-Seyler 367, 67-73, 1986  
A:Title: Studies on cytochrome-c oxidase, XIII. Amino-acid sequence of the small membra  
A:Reference number: A25879; MUID:86159303; PMID:3065725  
A:Accession: A25879  
A:Molecule type: protein  
A:Residues: 22-72, 'H', '74-75', 'KK' <MEI>  
A:Experimental source: heart  
R:Anthony, G.; Stroth, A.; Lottepeich, F.; Kadenbach, B.  
FEBS Lett. 277, 97-100, 1990  
A:Title: Different isoforms of cytochrome c oxidase are expressed in bovine smooth musc  
A:Reference number: S13001; MUID:91099535; PMID:2176624  
A:Accession: S13002  
A:Molecule type: protein  
A:Residues: 22-27 <ANT>  
A:Experimental source: smooth muscle  
R:Tsuchihara, T.; Aoyama, H.; Yamashita, E.; Tomizaki, T.; Yamaguchi, H.; Shinzawa-Itch,  
submitted to the Brookhaven Protein Data Bank, April 1996  
A:Reference number: A67451; PDB:1OCC  
A:Contents: annotation; X-ray crystallography, 2.8 angstrom, residues 22-77  
R:Tsuchihara, T.; Aoyama, H.; Yamashita, E.; Tomizaki, T.; Yamaguchi, H.; Shinzawa-Itch,  
Science 272, 1136-1144, 1996  
A:Title: The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 ang  
A:Reference number: A57981; MUID:96216288; PMID:8638158  
A:Contents: annotation; X-ray crystallography, 2.8 angstroms  
C:Genetics:  
A:Gene: COX7AH  
A:Insertions: 5/3; 34/3; 63/1  
C:Complex: part of a 13 chain complex, spanning the inner mitochondrial membrane and con  
(see PIR:OGB0), Vb (see PIR:OGB06a), VIIa (see PIR:OGB06), VIIb (see PIR:OGB07), VIC (se  
mers within the mitochondrial inner-membrane  
C:Function:  
A:Description: the cytochrome-c oxidase complex catalyzes the oxidation of four molecu  
ns from the mitochondrial matrix producing two molecules of water and lowering the conc.  
A:Pathway: oxidative phosphorylation; respiratory chain  
A>Note: the role of chain VIIa is not clear  
C:Superfamily: mammalian cytochrome-c oxidase chain VIIa  
C:Keywords: cardiac muscle; electron transfer; heart; membrane-associated complex; mito  
rane protein  
F:1-21/Domain: transit peptide (mitochondrion) #status predicted <TNP>  
F:22-80/Product: cytochrome-c oxidase chain VIIa #status experimental <MAT>  
F:22-46/Domain: mitochondrial matrix #status experimental <MM1>  
F:47-78/Domain: transmembrane helix #status experimental <TR01>  
F:79-80/Domain: intracrystal #status experimental <ITC1>

Query Match 78.0%; Score 32; DB 1; Length 80;  
Best Local Similarity 85.7%; Pred. No. 15;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

3 PVHLKRG 9  
| | | | |  
Db 40 PQHLKRG 46

Search completed: December 3, 2003, 15:53:26  
Job time: 3.63265 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 15:43:05 ; Search time 1.59184 Seconds  
(without alignments)  
265.882 Million cell updates/sec

Title: US-09-788-308d-1

Perfect score: 41

Sequence: 1 XPRVHLKRG 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	85.4	900	1	SUI2_DROME
2	34	82.9	367	1	Y502_MYCLE
3	34	82.9	467	1	SIL5_MOUSE
4	34	82.9	847	1	RSQ2_RAT
5	34	82.9	848	1	RSQ2_MOUSE
6	34	82.9	849	1	RSQ2_HUMAN
7	33	80.5	98	1	RL21_THERVO
8	33	80.5	114	1	CO7R_HUMAN
9	33	80.5	168	1	PSPC_RABIT
10	33	80.5	191	1	PSPC_MACMU
11	33	80.5	193	1	PSPC_MOUSE
12	33	80.5	194	1	PSPC_RAT
13	33	80.5	197	1	PSPC_HUMAN
14	33	80.5	294	1	HAL1_YEAST
15	33	80.5	505	1	DHAB_ORISA
16	33	80.5	695	1	RHNI_MOUSE
17	33	80.5	733	1	CANI_MOUSE
18	33	80.5	713	1	CANI_RAT
19	33	80.5	714	1	CANI_PIG
20	32	78.0	29	1	COXK_SHEEP
21	32	78.0	80	1	COXK_BOVIN
22	32	78.0	80	1	COXK_MOUSE
23	32	78.0	80	1	COXK_PIG
24	32	78.0	83	1	COXJ_BOVIN
25	32	78.0	83	1	COXJ_MOUSE
26	32	78.0	83	1	COXJ_RAT
27	32	78.0	507	1	MKR3_HUMAN
28	32	78.0	511	1	KPIC_SOYBN
29	32	78.0	628	1	GIDA_NEIMA
30	32	78.0	628	1	GIDA_NEIMA
31	32	78.0	1042	1	SVI_BORBU
32	32	78.0	1513	1	MUC2_RAT
33	32	78.0	2314	1	AKA6_RAT

34	32	78.0	3343	1	YOG7_CAEEL	P14616 caenorhabdi
35	32	78.0	5179	1	MUC2_HUMAN	Q02817 homo sapien
36	31	75.6	246	1	XPO7_MOUSE	Q9CTR8 mus musculu
37	31	75.6	346	1	Y1B9_YEAST	P40546 saccharomyc
38	31	75.6	472	1	DNAB_BACSV	P07908 bacillus su
39	31	75.6	489	1	Y092_METJA	Q57557 methanococc
40	31	75.6	574	1	Y091_YEAST	P40360 saccharomyc
41	31	75.6	643	1	RHNI_MOUSE	Q61085 mus musculu
42	31	75.6	736	1	VMI_REOVD	P12418 reovirus (t
43	31	75.6	962	1	XPO7_HUMAN	Q00335 reovirus (t
44	31	75.6	98	1	RL21_THERAC	Q43592 homo sapien
45	30	73.2				Q9H18 thermoplasm

## ALIGNMENTS

RESULT 1  
SUI2\_DROME STANDARD: PRT: 900 AA.  
AC Q9NUG9; Q8T9D8; Q9VW55;  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DE Polycomb protein Su(z)12 (Suppressor 12 of zeste protein).  
GN SU(Z)12 OR CG8013.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1), CHARACTERIZATION, AND MUTANT  
RP SU(Z)12-3.  
RX MEDLINE=21430867; PubMed=11546753;  
RA Birve A., Sengupta A.K., Beuchle D., Larsson J., Kennison J.A.,  
RA Rasmussen-Leestander A., Mueller J.,  
RT "Su(z)12, a novel Drosophila Polycomb group gene that is conserved in  
RT vertebrates and plants."  
RL Development 128:3371-3379(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkely;  
RX MEDLINE=20196006; PubMed=10711132;  
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfennoch C., Baldwin D.,  
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahler C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostettl M., Houson K.A., Howland T.J., Wei M.-H., Ilegem C.,  
RA Jialili M., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Polard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spiedling A.C., Stapleton M., Strong R., Sun E.,  
RA Svitek R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Massarom D.A., Weinstein G.M., Weisenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195 (2000).  
(3)  
RN REVISIONS, AND ALTERNATIVE SPLICING.  
RC STRAIN=Berkeley;  
RA MEDLINE=22426069; PubMed=12537572;  
RA Miura S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hirdocky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Bernan B.P.,  
RA Betencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Hartley N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
(4)  
RN SEQUENCE FROM N.A. (ISOFORM 2).  
RC STRAIN=Berkeley; TISSUE=Embryo;  
RA MEDLINE=22426066; PubMed=12537569;  
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
RA George R.A., Guatin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,  
RA Rubin G.M., Celniker S.E.;  
RT "A *Drosophila* full-length cDNA resource.";  
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
(5)  
RN IDENTIFICATION IN A BSC/E(Z) COMPLEX WITH E(Z); CAP1 AND ESC,  
RP AND METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.  
RX MEDLINE=22296673; PubMed=12408863;  
RA Czeizem B., Meli R., McCabe D., Seitz V., Imhof A., Pirodda V.;  
RT *Drosophila* enhancer of Zeste/ESC complexes have a histone H3  
RT methyltransferase activity that marks chromosomal Polycomb sites.";  
RL Cell 111:185-196(2002).  
(6)  
RN IDENTIFICATION IN A BSC/E(Z) COMPLEX WITH E(Z); CAP1 AND ESC, AND  
RP METHYLTRANSFERASE ACTIVITY OF THE COMPLEX.  
RX MEDLINE=22296674; PubMed=12408864;  
RA Mueller J., Hart C.M., Francis N.J., Vargas M.L., Sengupta A.,  
RA Wild B., Miller E.L., O'Connor M.B., Kingston R.E., Simon J.A.;  
RT "Histone methyltransferase activity of a *Drosophila* Polycomb group  
RT repressor complex.";  
RL Cell 111:197-208(2002).  
(7)  
RN FUNCTION: Polycomb group (PcG) protein. PcG proteins act by  
RN forming multiprotein complexes, which are required to maintain the  
RN transcriptionally repressive state of homeotic genes throughout  
RN development. PcG proteins are not required to initiate repression,  
RN but to maintain it during later stages of development. They  
RN probably act via the methylation of histones, rendering chromatin  
RN heritably changed in its expressibility. Component of the Bsc/E(z)  
RN complex, which methylates Lys-9 and Lys-27 residues of histone H3.  
RN Despite the presence of a zinc-finger, it does not bind directly  
RN to DNA, the Bsc/E(z) complex being probably recruited to DNA by  
RN Pho. The Bsc/E(z) complex is necessary but not sufficient to  
RN recruit a functional PcG repressive complex that represses target  
RN genes, suggesting that the recruitment of the distinct PRC1  
RN complex is also required to allow a subsequent repression.  
(8)  
RN SUBUNIT: Component of the Bsc/E(z) complex, composed of Bsc, E(z),  
RN Su(z)12, Rpd3, Caf1 and probably Pho. This complex is distinct  
RN from the PRC1 complex, which contains many other PcG proteins like  
RN Pc, Ph, Psc, Su(z)2. The two complexes however cooperate and  
RN interact together during the first 3 hours of development to  
RN establish PcG silencing.  
(9)  
RN SUBCELLULAR LOCATION: Nuclear.  
(10)  
RN ALTERNATIVE PRODUCTS:  
RN Event=Alternative splicing; Named isoforms=2;  
RN Name=1; Synonyms=B;

CC IsoId=Q9NU99-1; Sequence=displayed;  
CC Name=2; Synonyms=A;  
CC IsoId=Q9NU99-2; Sequence=VSP\_007033; VSP\_007034;  
CC Note=No experiment confirmation available;  
CC -1- SIMILARITY: Belongs to the VSPs (VRN2-EMF2-F152-SU(2)12) family.  
CC -1- SIMILARITY: Contains 1 C2H2-type zinc finger.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL, AF194047; AAF73149.1; -  
CC EMBL, AE003515; AAF49094.2; -  
CC EMBL, AE003515; AAN1641.1; -  
CC EMBL, AY069809; AAL39954.1; -  
CC Flybase: FBgn020887; Su(z)12.  
CC InterPro: IPR007087; Znf C2H2.  
CC SMART, SM00355; Znf C2H2\_1.  
CC PROSITE, PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
CC Transcription regulation; Repressor; Developmental protein;  
CC Nuclear protein; Metal-binding; Zinc; Zinc-finger;  
CC Alternative splicing.  
CC ZN FING 411 434  
CC DOMAIN 527 603  
CC DOMAIN 355 366  
CC DOMAIN 699 879  
CC VARSPLIC 806 855  
CC FT FT  
CC FT VARSPLIC 856 900  
CC FT VARSPLIC 856 900  
CC FT MUTAGEN 274 274  
CC FT FT  
CC SQ SEQUENCE 900 AA; 100104 MW; 53BAD83C49BC92F CRC64;  
  
CC Query Match 85.4%; Score 35; DB 1; Length 900;  
CC Best Local Similarity 85.7%; Pred. No. 14;  
CC Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
CC QY 3 PVLKRG 9  
CC Db 722 PVLKRG 728  
  
CC RESULT 2  
CC Y502 MYCLE  
CC ID Y502 MYCLE STANDARD; PRT; 367 AA.  
CC AC P54678; Q9CB58;  
CC DT 01-OCT-1996 (Rel. 34, Created)  
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)  
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
CC DE Hypothetical protein ML2427.  
CC GN ML2427 OR B2168\_C1.175.  
CC OS Mycobacterium leprae.  
CC OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
CC OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
CC OC NCBI\_TaxID=1769;  
CC RN 11)  
CC RP SEQUENCE FROM N.A.  
CC RA Smith D.R., Robison K.;  
CC RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC (2)  
CC RP SEQUENCE FROM N.A.  
CC RC STRAIN=TN;  
CC RX MEDLINE=21126732; PubMed=11234002;  
CC RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,  
CC RA Wheeler P.R., Honore N., Garnier J., Churcher C., Harris D.,  
CC RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,





```

FT MUTAGEN 432 432 Y->F: ABOLISHES BINDING TO PTPN6 AND
FT MUTAGEN 455 455 PTPN11.
FT CONFLICT 2 2 Y->F: REDUCES BINDING TO PTPN6.
FT CONFLICT 24 24 L -> M (IN REF. 2).
FT CONFLICT 102 102 V -> G (IN REF. 2).
FT CONFLICT 102 102 D -> E (IN REF. 2).
FT CONFLICT 234 234 T -> I (IN REF. 2).
SQ SEQUENCE 467 AA; 51889 MW; 54FD3E8950D2DE CRC64;

Query Match
Best Local Similarity 82.9%; Score 34; DB 1; Length 467;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 PVHLKRG 9
Db 312 PVHLKRG 318

RESULT 4
RSG2_RAT STANDARD; PRT; 847 AA.
AC 063713;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Ras GTPase-activating protein 2 (GAP1m).
GN RASA2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95021216; PubMed=7935405;
RA Maekawa M., Li S., Yamatsu A., Morishita T., Yokota K., Imai Y.,
RA Kohaka S., Nakamura S., Hattori S.;
RA "A novel mammalian Ras GTPase-activating protein which has
RT phospholipid-binding and Btk homology regions.";
RT Mol. Cell. Biol. 14:6879-6885(1994).
CC - FUNCTION: INHIBITORY REGULATOR OF THE RAS-CYCLIC AMP PATHWAY. MAY
CC BIND INOSITOL TETRAKISPHOSPHATE (IP4) AND PHOSPHOLIPIDS.
CC - SUBCELLULAR LOCATION: PLASMA MEMBRANE (POTENTIAL).
CC - TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHER EXPRESSION IN BRAIN,
CC PLACENTA, AND KIDNEY.
CC - SIMILARITY: Contains 2 C2 domains.
CC - SIMILARITY: Contains 1 BTK domain.
CC - SIMILARITY: Contains 1 Ras-GAP domain.
CC - SIMILARITY: Contains 1 Ras-GAP domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, D30734; BAA06398.1; -
CC PIR, A56039; A56039.
CC HSSP, P21707; IRSY.
CC InterPro, IPR001562; BTK.
CC InterPro, IPR000008; C2.
CC InterPro, IPR001849; PH.
CC InterPro, IPR001936; RasGAP.
CC Pfam, PF00779; BTK; 1.
CC Pfam, PF00168; C2; 2.
CC Pfam, PF00169; PH; 1.
CC Pfam, PF00616; RasGAP; 1.
CC PRINTS, PR00402; TECHTDOMAIN.
CC SMART, SM00107; BTK; 1.
CC SMART, SM00239; C2; 2.
CC SMART, SM00233; PH; 1.

```

```

DR SMART, SM00233; RasGAP; 1.
DR PROSITE, PS50003; PH DOMAIN; 1.
DR PROSITE, PS00499; C2 DOMAIN; 1; FALSE_NEG.
DR PROSITE, PS50004; C2 DOMAIN; 2.
DR PROSITE, PS00509; RAS_GTPASE_ACTIV_1; FALSE_NEG.
DR PROSITE, PS50018; RAS_GTPASE_ACTIV_2; 1.
KW GTPase activation; Repeat.
FT DOMAIN 24 121 C2 DOMAIN 1.
FT DOMAIN 165 272 C2 DOMAIN 2.
FT DOMAIN 355 549 RAS-GAP.
FT DOMAIN 603 704 PH.
FT DOMAIN 706 742 BTK.
SQ SEQUENCE 847 AA; 96373 MW; A7E88AF2DA74516B CRC64;

Query Match
Best Local Similarity 82.9%; Score 34; DB 1; Length 847;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 PVHLKRG 9
Db 602 PVHLKRG 608

RESULT 5
RSG2_MOUSE STANDARD; PRT; 848 AA.
AC P58069;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ras GTPase-activating protein 2 (GAP1m).
GN RASA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RC TISSUE=Brain;
RX MEDLINE=95324967; PubMed=8702543;
RA Fukuda M., Mikoshiba K.;
RA "Structure-function relationships of the mouse Gap1m: Determination of
RT the inositol 1,3,4,5-tetrakisphosphate-binding domain.";
RT J. Biol. Chem. 271:18838-18842(1996).
CC - FUNCTION: INHIBITORY REGULATOR OF THE RAS-CYCLIC AMP PATHWAY.
CC BINDS INOSITOL TETRAKISPHOSPHATE (IP4) AND PHOSPHOLIPIDS.
CC - SUBCELLULAR LOCATION: PLASMA MEMBRANE (POTENTIAL).
CC - SIMILARITY: Contains 2 C2 domains.
CC - SIMILARITY: Contains 1 BTK domain.
CC - SIMILARITY: Contains 1 BTK domain.
CC - SIMILARITY: Contains 1 Ras-GAP domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AB056433; BAB32975.1; -
CC HSSP, MGI:2149860; Rasea2.
CC MGD, MGI:2149860; Rasea2.
CC InterPro, IPR001562; BTK.
CC InterPro, IPR000008; C2.
CC InterPro, IPR001849; PH.
CC InterPro, IPR001936; RasGAP.
CC Pfam, PF00779; BTK; 1.
CC Pfam, PF00168; C2; 2.
CC Pfam, PF00169; PH; 1.
CC Pfam, PF00616; RasGAP; 1.
CC PRINTS, PR00402; TECHTDOMAIN.

```

```

DR SMART; SM00107; BTK; 1.
DR SMART; SM00239; C2; 2.
DR SMART; SM00233; PH; 1.
DR SMART; SM00323; RaasGAP; 1.
DR PROSITE; PSS00003; PH_DOMAIN; 1.
DR PROSITE; PSS00499; C2_DOMAIN_1; FALSE_NEG.
DR PROSITE; PSS00004; C2_DOMAIN_2; 2.
DR PROSITE; PSS00509; RAS_GTPASE_ACTIV_1; FALSE_NEG.
DR PROSITE; PSS00018; RAS_GTPASE_ACTIV_2; 1.
KM GTPase activation; Repeat.
FT DOMAIN 25 122 C2 DOMAIN 1.
FT DOMAIN 166 273 C2 DOMAIN 2.
FT DOMAIN 356 550 RAS-GAP.
FT DOMAIN 604 705 PH.
FT DOMAIN 707 743 BTK.
FT DOMAIN 1 22 ALA-RICH.
FT MUTAGEN 629 629 R->C: GREATLY REDUCED BINDING TO IP4 AND
TO PHOSPHOLIPIDS.
FT MUTAGEN 627 629 KKR->QQQ: NO BINDING TO IP4, REDUCED
BINDING TO PHOSPHOLIPIDS.
SQ SEQUENCE 848 AA; 96472 MW; 8F1DCD5D59108B5 CRC64;

Query Match 82.9%; Score 34; DB 1; Length 848;
Best Local Similarity 85.7%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 PVALKRG 9
DB 603 PVALKRG 609

RESULT 6
RSG2 HUMAN STANDARD; PRT; 849 AA.
AC Q15283; Q00695; Q15284; Q92594; Q99577; Q91EQ2;
ID 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Ras GTPase-activating protein 2 (GAP1m).
GN RAS2 OR RASGAP OR GAP1m.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=97074668; PubMed=8917095;
RA Kobayashi M., Masui T., Kusuda J., Kameoka Y., Hashimoto K.,
RA Iwashita S.;
RT "Human rasGTPase-activating protein (human counterpart of GAP1m):
RT sequence of the cDNA, primary structure of the protein, production and
RT chromosomal localization.";
RL Gene 175:173-177(1996).
RN SEQUENCE FROM N.A.
RX MEDLINE=97001173; PubMed=8812506;
RA Li S., Satoh H., Matanabe T., Nakamura S., Hattori S.;
RT "cDNA cloning and chromosomal mapping of a novel human GAP (GAP1m), a
RT GTPase-activating protein of Ras.";
RL Genomics 35:625-627(1996).
[3]
RN SEQUENCE FROM N.A.
RX MEDLINE=98044291; PubMed=9382842;
RA Lockyer P.J., Bottomley J.R., Reynolds J.S., McNulty T.J.,
RA Venkateswarlu K., Potter B.V.L., Dempsey C.E., Culen P.J.;
RT "Distinct subcellular localisations of the putative inositol 1,3,4,5-
RT tetraakisphosphate receptors GAP1(IP4BP) and GAP1m result from the
RT GAP1(IP4BP) PH domain directing plasma membrane targeting.";
RL Curr. Biol. 7:1007-1010(1997).
[4]
RN SEQUENCE FROM N.A.

```

```

RC TISSUE=Blood;
RA Lockyer P.J.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION. INHIBITORY REGULATOR OF THE RAS CYCLIC AMP PATHWAY.
CC BINDS INOSITOL TETRAKISPHOSPHATE (IP4).
CC -1- SUBCELLULAR LOCATION. PERINUCLEAR AND CYTOPLASMIC.
CC -1- SIMILARITY. Contains 2 C2 domains.
CC -1- SIMILARITY. Contains 1 PH domain.
CC -1- SIMILARITY. Contains 1 BTK domain.
CC -1- SIMILARITY. Contains 1 Ras-GAP domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D78155; BA11230.1; -
DR EMBL; D78156; BA11231.1; -
DR EMBL; D82880; BA11621.1; -
DR EMBL; D82881; BA11622.1; -
DR EMBL; AF115573; AAD09821.1; -
DR PIR; JC5047; JC5047.
DR HSSP; Q06187; IBTK.
DR Genew; HGNC:9872; RAS2.
DR MIM; 601589; .
DR GO; GO:0005099; F:RAS GTPase activator activity; TAS.
DR Interpro; IPR001562; BTK.
DR Interpro; IPR000008; C2.
DR Interpro; IPR001849; PH.
DR Interpro; IPR001936; RaasGAP.
DR Pfam; PF00779; BTK; 1.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00616; RaasGAP; 1.
DR PRINTS; PRO0402; TERCPTDOMAIN.
DR SMART; SM00107; BTK; 1.
DR SMART; SM00239; C2; 2.
DR SMART; SM00233; PH; 1.
DR SMART; SM00323; RaasGAP; 1.
DR PROSITE; PSS00003; PH_DOMAIN; 1.
DR PROSITE; PSS00499; C2_DOMAIN_1; FALSE_NEG.
DR PROSITE; PSS00004; C2_DOMAIN_2; 2.
DR PROSITE; PSS00509; RAS_GTPASE_ACTIV_1; FALSE_NEG.
DR PROSITE; PSS00018; RAS_GTPASE_ACTIV_2; 1.
KM GTPase activation; Repeat.
FT DOMAIN 25 122 C2 DOMAIN 1.
FT DOMAIN 166 273 C2 DOMAIN 2.
FT DOMAIN 356 550 RAS-GAP.
FT DOMAIN 604 705 PH.
FT DOMAIN 707 743 BTK.
FT DOMAIN 1 26 ALA-RICH.
FT CONFLICT 216 216 T -> A (IN REF. 1).
FT CONFLICT 645 645 G -> GS (IN REF. 1).
FT CONFLICT 645 645 G -> EFIR (IN REF. 2).
SQ SEQUENCE 849 AA; 96526 MW; A4B491DFF5C4CB76 CRC64;

Query Match 82.9%; Score 34; DB 1; Length 849;
Best Local Similarity 85.7%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 PVALKRG 9
DB 603 PVALKRG 609

RESULT 7
RL21 THEVO STANDARD; PRT; 98 AA.
ID RL21 THEVO
AC Q97BZ1;
DT 28-FEB-2003 (Rel. 41, Created)

```

DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 50S ribosomal protein L21e.  
GN RPL21E OR TV0314 OR TVG0322585.  
OS Thermoplasma volcanium.  
OC Archaea; Euryarchaeota; Thermoplasmales;  
OC Thermoplasmales; Thermoplasma.  
ON NCBI\_TaxID=50339;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GS1 / DSM 4299 / JCM 9571;  
RX MEDLINE=20570466; PubMed=11121031;  
RA Kawashima T., Ameno N., Koike H., Makino S.-I., Higuchi S.,  
RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,  
RA Numoshita T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.,  
RT "Archaeal adaptation to higher temperatures revealed by genomic  
RT sequence of Thermoplasma volcanium."  
RL Proc. Natl. Acad. Sci. U.S.A. 97:1425-1426(2000).  
CC - SIMILARITY: BELONGS TO THE L21E FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AP000997; BAB59456.1; -  
DR HAMAP: MF\_00369; -; 1.  
DR InterPro: IPR001147; Ribosomal\_L21e.  
DR Pfam: PF01157; Ribosomal\_L21e; 1.  
DR PROSITE: PS01171; RIBOSOMAL\_L21e; 1.  
KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 98 AA; 10915 MW; 825AFED1E3DE2C9C CRC64;  
OY 3 PVTLRK 8  
Db 90 PVTLRK 95  
Query Match 80.5%; Score 33; DB 1; Length 98;  
Best Local Similarity 100.0%; Pred. No. 4.3;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 3 PVTLRK 8  
Db 90 PVTLRK 95  
RESULT 8  
CO7R HUMAN  
ID CO7R HUMAN STANDARD; PRT; 114 AA.  
AC O14548; Q9P118;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Cytochrome c oxidase subunit VIIA-related protein, mitochondrial  
DE precursor (COX7A-related protein) (EBI).  
GN COX7A2L OR COX7RP OR COX7AR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
ON NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=98078700; PubMed=9418891;  
RA Watanabe T., Inoue S., Hiroi H., Orito A., Kawashima H., Muramatsu M.,  
RT "Isolation of estrogen-responsive genes with a Cpg island library."  
RL Mol. Cell. Biol. 18:442-449(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99268136; PubMed=10335655;  
RA Schmidt T.R., Goodman M., Grossman L.I.,  
RT "Molecular evolution of the COX7A gene family in primates."  
RL Mol. Biol. Evol. 16:619-626(1999).  
RN [3]

RP SEQUENCE FROM N.A.  
RX MEDLINE=21090505; PubMed=1156535;  
RA Lee N., Daly M.J., Delmonte T., Lander E.S., Xu F., Hudson T.J.,  
RA Mitchell G.A., Morin C.C., Robinson B.H., Roux J.D.,  
RT "A genome-wide linkage-disequilibrium scan localizes the  
RT Saguena-y-lac-Saint-Jean cytochrome oxidase deficiency to 2p16."  
RL Am. J. Hum. Genet. 68:397-409(2001).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung, and Prostate;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heslop F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshlyuk S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butlerfield V.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Maitra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences."  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC - FUNCTION: MAY BE A REGULATORY SUBUNIT OF CYTOCHROME C OXIDASE THAT  
CC MEDIATES THE HIGHER LEVEL OF ENERGY PRODUCTION IN TARGET CELLS BY  
CC ESTROGEN.  
CC -----  
CC - SUBCELLULAR LOCATION: Mitochondrial (by similarity).  
CC -----  
CC - INDUCTION: By estrogen.  
CC -----  
CC - SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE VIIA FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AB007618; BAA22571.1; -  
DR EMBL: AF127788; AAF72746.1; -  
DR EMBL: AY007643; AAG32129.1; -  
DR EMBL: BC005251; AAH05251.1; -  
DR EMBL: BC007095; AAH07095.1; -  
DR Gene: HGNC:2289; COX7A2L.  
DR MIM: 605771;  
DR GO: GO:004129; F:cytochrome c oxidase activity; TAS.  
DR InterPro: IPR003177; COX7A.  
DR Pfam: PF02238; COX7A; 1.  
DR ProDom: PD006023; COX7A; 1.  
KW Mitochondrion; Transist peptide.  
FT TRANSIT 1 55  
FT CHAIN 56 114  
FT CYTOCHROME C OXIDASE SUBUNIT VIIA-RELATED  
FT PROTEIN.  
FT CONFLICT 107 107 M -> N (IN REF. 1).  
SQ SEQUENCE 114 AA; 12615 MW; 67BACABEEA0F52B CRC64;  
OY 3 PVTLRK 9  
Db 74 PVTLRK 80  
Query Match 80.5%; Score 33; DB 1; Length 114;  
Best Local Similarity 85.7%; Pred. No. 4.9;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 9
PSPC RABIT STANDARD; PRT: 188 AA.
ID P23358;
AC 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pulmonary surfactant-associated protein C precursor (SP-C) (Pulmonary
  surfactant-associated proteolipid SPL(Val1)).
GN SFTPC OR SFTPD2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
[1]
SEQUENCE FROM N.A.
RA STRAIN=New Zealand white; TISSUE=Lung;
RX MEDLINE=9235123; PubMed=1643107;
RA Connelly I., Possmayer F.;
RT "CDNA sequence and alternative mRNA splicing of surfactant-associated
  protein C (SP-C) in rabbit lung."
RL Biochim. Biophys. Acta 1127:199-207(1992).
RN [2]
SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=93118799; PubMed=1335697;
RA Boggaram V., Margana R.K.;
RT "Rabbit surfactant protein C: cDNA cloning and regulation of
  alternatively spliced surfactant protein C mRNAs."
RL Am. J. Physiol. 263:L634-L644(1992).
RN [3]
SEQUENCE FROM N.A.
RA Margana R.K., Boggaram V.;
RT "Cloning, sequence and characterization of the rabbit surfactant
  protein C gene."
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
SEQUENCE OF 24-188 FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=93105936; PubMed=1468410;
RA Durham P.L., Nanthakumar E.J., Snyder J.M.;
RT "Developmental regulation of surfactant-associated proteins in rabbit
  fetal lung in vivo."
RL Exp. Lung Res. 18:775-793(1992).
RN [5]
SEQUENCE OF 24-37.
RX MEDLINE=91200266; PubMed=2015882;
RA Johansson U., Persson P., Loewenadler B., Robertson B., Joernvall H.,
  Curstedt T.;
RT "Canine hydrophobic surfactant polypeptide SP-C. A lipopeptide with
  one cholesterol-linked palmitoyl group."
RL FEBS Lett. 281:119-122(1991).
RN [6]
FUNCTION: PULMONARY SURFACTANT ASSOCIATED PROTEINS PROMOTE
  ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-
  LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES.
CC [1] SUBCELLULAR LOCATION: Extracellular.
CC [1] MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
  PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
  CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
  HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC [7]
This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
  or send an email to license@ebi.ac.uk).
```

```

DR EMBL; S51098; AAB24576.2; -
DR PIR; A56766; LNRBC1.
DR InterPro; IPR001729; Pulm_surfact_AP.
DR Pfam; PF04089; BRICHOS; 1.
DR ProDom; PD009591; Pulm_surfact_AP; 1.
DR SMART; SM00019; SP_P; 1.
DR PROSITE; PS00341; SURFACT_PALMITOYL; 1.
KW Surfact film; Gaseous exchange; Lipoprotein; Palmitate.
FT PROPEP 1 23
FT CHAIN 24 58
FT FT C.
FT PROPEP 59 188
FT LIPID 28 28
FT LIPID 29 29
FT CONFLICT 115 115
FT CONFLICT 153 153
FT CONFLICT 159 159
FT CONFLICT 161 161
FT CONFLICT 186 186
SQ SEQUENCE 188 AA; 19836 MW; F622EBA933786F78 CRC64;

Query Match 80.5%; Score 33; DB 1; Length 188;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 PVHLKR 8
Db 30 PVHLKR 35

RESULT 10
PSPC MACMU STANDARD; PRT: 191 AA.
ID P55152;
AC 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pulmonary surfactant-associated protein C precursor (SP-C) (Pulmonary
  surfactant-associated proteolipid SPL(Val1)).
GN SFTPC OR SFTPD2.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
  OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9544;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA An G., Luo G., Zhao Y., Plopper C., Wu R.;
RT Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
FUNCTION: PULMONARY SURFACTANT ASSOCIATED PROTEINS PROMOTE
  ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-
  LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES.
CC [1] SUBCELLULAR LOCATION: Extracellular.
CC [1] MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
  PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
  CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
  HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC [7]
This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
  or send an email to license@ebi.ac.uk).
```



Query Match 80.5%; Score 33; DB 1; Length 194;  
 Best Local Similarity 100.0%; Pred. No. 8.3;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKR 8  
 |||||  
 DB 30 PVHLKR 35

RESULT 13  
 PSPC\_HUMAN STANDARD; PRT; 197 AA.  
 ID\_PSPC\_HUMAN P1687; Q12793;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Pulmonary surfactant-associated protein C precursor (SP-C) (SP5)  
 GN SFTPC OR SFTP2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-42.  
 RX MEDLINE=88087156; PubMed=335510;  
 RA Glaeser S.W., Korfhagen T.R., Weaver T.E., Clark J.C.,  
 RA Pilot-Matias T., Meuth J., Fox J.L., Whitsett J.A.;  
 RT "cDNA, deduced polypeptide structure and chromosomal assignment of  
 human pulmonary surfactant proteolipid, SPL(pval).";  
 RL J. Biol. Chem. 263:9-12(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8827133; PubMed=2839484;  
 RA Glaeser S.W., Korfhagen T.R., Perme C.M., Pilot-Matias T.J.,  
 RA Kuster S.E., Whitsett J.A.;  
 RT "Two SP-C genes encoding human pulmonary surfactant proteolipid.";  
 RL J. Biol. Chem. 263:10326-10331(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8606508; PubMed=3479771;  
 RA Watt R.G., Hawgood S., Buckley D.I., Crisp T.M., Schilling J.,  
 RA Benson B.J., Ballard P.L., Clements J.A., White R.T.;  
 RT "Low molecular weight human pulmonary surfactant protein (SP5):  
 RT isolation, characterization, and cDNA and amino acid sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:7915-7919(1987).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94237133; PubMed=8181452;  
 RA Hatzis D., Delter G., demello D.E., Floros J.;  
 RT "Human surfactant protein-C: genetic homogeneity and expression in  
 RT RDS: comparison with other species.";  
 RL Exp. Lung Res. 20:57-72(1994).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.J.,  
 RA Raha S.S., Lottuano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Wooley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smajlus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [6]  
 RP SEQUENCE OF 24-58.  
 RX MEDLINE=88211876; PubMed=3366248;  
 RA Johanson J., Joernvall H., Eklund A., Christensen N., Robertson B.,  
 RA Cusack T.;  
 RT "Hydrophobic 3.7 kDa surfactant polypeptide: structural  
 RT characterization of the human and bovine forms.";  
 RL FEBS Lett. 232:61-64(1988).  
 RN [7]  
 RP PALMITOYLATION  
 RX MEDLINE=90222154; PubMed=2326260;  
 RA Cusack T., Johanson J., Persson P., Eklund A., Robertson B.,  
 RA Loewenadler B., Joernvall H.;  
 RT "Hydrophobic surfactant-associated polypeptides: SP-C is a  
 RT lipopeptide with two palmitoylated cysteine residues, whereas SP-B  
 RT lacks covalently linked fatty acyl groups.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2985-2989(1990).  
 CC -1- FUNCTION: PULMONARY SURFACTANT ASSOCIATED PROTEINS PROMOTE  
 CC ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-  
 CC LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=P1686-1; Sequence=Displayed;  
 CC Name=Cl;  
 CC IsoId=P1686-2; Sequence=VSP\_006311;  
 CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%  
 CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,  
 CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL  
 CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).  
 CC CC  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib.ch](mailto:license@sib.ch)).  
 CC -----  
 DR EMBL; J03517; AAA3634.1; -;  
 DR EMBL; J03890; AAC32022.1; -;  
 DR EMBL; J03890; AAC32023.1; -;  
 DR EMBL; J03553; AAA3631.1; -;  
 DR EMBL; U02948; AAB60332.1; -;  
 DR EMBL; BC005913; AAH05913.1; -;  
 DR PIR; A28801; LNHUC.  
 DR Genew; HGNC:10802; SFTPC.  
 DR MIM; 178620; -;  
 DR InterPro; IPR001729; Pulm surfact AP.  
 DR Pfam; PF04089; BRICHOS; 1;  
 DR ProDom; PD008591; Pulm\_surfact\_AP; 1.  
 DR SMART; SM00019; SP\_P; 1.  
 DR PROSITE; PS00341; SURFACT\_PALMITOYL; 1.  
 DR Surface film; Gaseous exchange; Lipoprotein; Palmitate; Polymorphism;  
 KW Alternative splicing.  
 FT PROPEP 1 23  
 FT CHAIN 24 58  
 FT C.  
 FT PALMITATE.  
 FT LIPID 28 28  
 FT LIPID 29 29  
 FT VASAPLIC 146 151  
 FT VARIANT 138 138  
 FT N -> T  
 FT /FTId=VAR\_007453.  
 FT P -> PQC (IN REF. 4).  
 FT L -> S (IN REF. 4).  
 FT CONFLICT 14 14  
 FT CONFLICT 45 45

```

FT CONFLICT 65 67 TEM -> FPO (1N REF. 4)
FT CONFLICT 186 186 N -> S (1N REF. 4)
SQ SEQUENCE 197 AA; 21053 MW; C26A21E33C60AA78 CRC64;

Query Match
Best Local Similarity 80.5%; Score 33; DB 1; Length 197;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKR 8
DB 30 PVHLKR 35

RESULT 14
HALI_YEAST STANDARD; PRT; 294 AA.
ID HALI_YEAST
AC Q01766
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Halotolerance protein HALI.
GN HALI OR YPR005C OR YP9723.05C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSBY939.
RX MEDLINE=92371421; PubMed=1505513;
RA Gaxiola R., de Larrinoa I.F., Villalba J.M., Serrano R.;
RT "A novel and conserved salt-induced protein is an important
RT determinant of salt tolerance in yeast."
RL EMBL J. 11:3157-3164(1992).

[2]
RP SEQUENCE FROM N.A.
RC STRAIN=DS288C / AB972;
RX MEDLINE=9713371; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Heblung U., Heumann K., Hilbert H., Hillier L.,
RA Hunnicutt-Smith S., Hyman R., Johnson M., Kaiman S., Kline K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oetner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharte M.,
RA Scherrens B., Schramm S., Schroeder M., Sidcu A.M., Tectelin H.,
RA Uresterazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.W., Wamburt R., Wang Y., Wedler E., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI."
RL Nature 387:103-105(1997).
CC -1- FUNCTION: INVOLVED IN SALT TOLERANCE.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- INDUCTION: By salt stress.

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U31900; AAA97584.1; -
CC EMBL; X67559; CAA47858.1; -
CC EMBL; Z48951; CAA88783.1; -
CC EMBL; Z71255; CAA95045.1; -
CC PIR; S23561; S23561.

```

```

DR SGD; S0006209; HALI.
DR GO; GO:0005737; Cytoplasm; IDA.
DR GO; GO:0009651; P:salinity response; IGI.
DR GO; GO:0006367; P:transcription initiation from pol II promoter; IMP.
SQ SEQUENCE 294 AA; 32994 MW; 85C4F35ADCB78BF CRC64;

Query Match
Best Local Similarity 85.7%; Score 33; DB 1; Length 294;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 PVHLKR 9
DB 112 PVHLKR 118

RESULT 15
DHAB_ORYSA STANDARD; PRT; 505 AA.
ID DHAB_ORYSA
AC Q24174;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Betaine-aldehyde dehydrogenase (EC 1.2.1.8) (BADH).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoideae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RX MEDLINE=97336302; PubMed=9193078;
RA Nakamura T., Yokota S., Muramoto Y., Tsutsui K., Oguri Y., Fukui K.,
RA Takebe T.;
RT "Expression of a betaine aldehyde dehydrogenase gene in rice, a
RT glycinebetaine nonaccumulator, and possible localization of its
RT protein in peroxisomes."
RL Plant J. 11:1115-1120(1997).
CC -1- CATALYTIC ACTIVITY: Betaine aldehyde + NAD(+) + H(2)O = betaine +
CC NADH.
CC -1- PATHWAY: Betaine biosynthesis; last step.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Peroxisomal.
CC -1- SIMILARITY: Belongs to the aldehyde dehydrogenase family.

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AB001348; BAA21098.1; -
CC PIR; T03394; T03394.
DR HSSP; P05091; 1CW3.
DR Gramene; Q24174; -
DR InterPro; IPR02086; Aldehyde_dehydr.
DR Pfam; PF00171; aldehyd. 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1
KW Oxidoreductase; NAD; Peroxisome.
NP BIND 240 245
FT ACT_SITE 262 262 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 286 286 BY SIMILARITY.
FT ACT_SITE 296 296 BY SIMILARITY.
FT SITE 503 505 MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ SEQUENCE 505 AA; 54647 MW; 85EFAA2B059A8081 CRC64;

Query Match
Best Local Similarity 85.7%; Score 33; DB 1; Length 505;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 PVHLKR 9

```

Db 368 POH,KRG 374

Search completed: December 3, 2003, 15:50:31  
Job time : 1.59184 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using 6w model

Run on: December 3, 2003, 15:44:20 ; Search time 6.42857 Seconds  
(without alignments)  
361.274 Million cell updates/sec

Title: US-09-788-308d-1

Perfect score: 41

Sequence: 1 XXPHVLRG 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	92.7	113	11	Q99KD6
2	36	87.8	373	16	Q8ZPN1
3	36	87.8	373	16	Q8ZPN1
4	36	87.8	792	16	Q8XKL7
5	35	85.4	221	16	Q98ND1
6	35	85.4	319	10	Q9XF81
7	35	85.4	855	5	Q9VM55
8	35	85.4	900	5	Q9NIG9
9	34	82.9	170	11	Q9IVC0
10	34	82.9	170	11	Q9QY93
11	34	82.9	492	17	Q27878
12	34	82.9	492	17	Q27878
13	33	80.5	86	11	Q99MI3
14	33	80.5	231	5	Q81IK9
15	33	80.5	251	10	Q94F79
16	33	80.5	258	10	Q94F78

17	33	80.5	323	5	Q9VSW5
18	33	80.5	337	17	Q8ZPF5
19	33	80.5	494	10	Q43829
20	33	80.5	650	6	Q9NOM7
21	33	80.5	670	4	Q8TAV1
22	33	80.5	695	4	Q8TCX5
23	33	80.5	713	11	Q8C2J1
24	32	78.0	162	11	P70598
25	32	78.0	181	4	Q00243
26	32	78.0	210	4	Q9UN50
27	32	78.0	210	4	Q14613
28	32	78.0	214	11	Q8J2X9
29	32	78.0	254	4	Q8N150
30	32	78.0	301	11	Q9Z2U5
31	32	78.0	510	10	Q8L7J4
32	32	78.0	511	10	Q8L7J5
33	32	78.0	718	5	Q9V594
34	32	78.0	751	5	Q8MKU2
35	32	78.0	778	13	Q98TP4
36	32	78.0	913	3	Q12151
37	32	78.0	1487	4	Q9TUM3
38	32	78.0	3337	5	Q9TWY4
39	31	75.6	101	17	Q9Y9A5
40	31	75.6	124	2	Q9KH43
41	31	75.6	139	2	Q54621
42	31	75.6	143	10	Q8LHW5
43	31	75.6	185	16	Q9P9R0
44	31	75.6	217	11	Q8BX14
45	31	75.6	247	2	Q85849

## ALIGNMENTS

RESULT 1  
Q99KD6 PRELIMINARY; PRT; 113 AA.  
AC Q99KD6;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Similar to silica-induced gene 81.  
GN COX7A2L.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC004715; AAH04715.1; -  
DR MGD; MGI:106015; Cox7a2l.  
DR InterPro; IPR003177; COX7a.  
DR Pfam; PF02238; COX7a; 1.  
SQ SEQUENCE 113 AA; 12562 MW; 1E2BFSDCBB6772D4 CRC64;

Query Match 92.7%; Score 38; DB 11; Length 113;  
Best Local Similarity 85.7%; Pred. No. 3.2;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKRG 9  
DB 73 PVHLKRG 79

RESULT 2  
ID Q8ZPN1 PRELIMINARY; PRT; 373 AA.  
AC Q8ZPN1;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Putative cytoplasmic protein.  
GN YDHH OR STM1446.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCB1\_Taxid=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Speth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
LT2.";  
RL Nature 413:852-856(2001).  
DR EMBL; AE008762; AAL20368.1; -  
DR InterPro; IPR005338; UPF0075.  
DR Pfam; PF03702; UPF0075; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 373 AA; 39735 MW; 043516A7BE1CE285 CRC64;  
  
Query Match 87.8%; Score 36; DB 16; Length 373;  
Best Local Similarity 85.7%; Pred. No. 26;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 PVLKRG 9  
| | | | | | | | | |  
Db 41 PVLKRG 47  
  
RESULT 3  
Q82600 PRELIMINARY; PRT; 373 AA.  
AC Q82600;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein STY1676.  
GN STY1676.  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCB1\_Taxid=601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,  
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
RA Cronin A., Davis P., Davies R.M., Dow L., White N., Farrar J.,  
RA Letwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,  
RA Ktogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrett B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
enterica serovar Typhi CT18.";  
RL Nature 413:848-852(2001).  
DR EMBL; AL627271; CAD01921.1; -  
DR InterPro; IPR005338; UPF0075.  
DR Pfam; PF03702; UPF0075; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 373 AA; 39795 MW; 6EF516A3EB575260 CRC64;  
  
Query Match 87.8%; Score 36; DB 16; Length 373;  
Best Local Similarity 85.7%; Pred. No. 26;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 PVLKRG 9  
| | | | | | | | | |  
Db 41 PVLKRG 47

Db 41 PVLKRG 47  
  
RESULT 4  
Q8XXL7 PRELIMINARY; PRT; 792 AA.  
AC Q8XXL7;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Probable xanthine dehydrogenase (subunit B) oxidoreductase protein  
DE (EC 1.1.1.204).  
GN XDH OR RSC2096 OR R803661.  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Ralstoniaceae; Ralstonia.  
OX NCB1\_Taxid=305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GM11000;  
RX MEDLINE=21681879; PubMed=11823852;  
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
RA Arlat M., Billault A., Brottier P., Camus J.C., Catolico L.,  
RA Chandler M., Cholsene N., Claudel-Renard C., Cunac S., Demange N.,  
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,  
RA Signier P., Thebaud P., Whalen M., Wincker P., Levy M.,  
RA Weisenbach J., Boucher C.A.;  
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";  
RL Nature 415:497-502(2002).  
DR EMBL; AL646068; CAD15803.1; -  
DR InterPro; IPR000674; Aldxan\_dh\_C;  
DR Pfam; PF03315; Ald\_xan\_dh\_C\_1;  
DR Pfam; PF02738; Ald\_xan\_dh\_C2\_1;  
KW Oxidoreductase; Complete proteome.  
SQ SEQUENCE 792 AA; 85409 MW; BF8BC5B209BEB49E CRC64;  
  
Query Match 87.8%; Score 36; DB 16; Length 792;  
Best Local Similarity 85.7%; Pred. No. 54;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 3 PVLKRG 9  
| | | | | | | | | |  
Db 166 PVLKRG 172  
  
RESULT 5  
Q98ND1 PRELIMINARY; PRT; 221 AA.  
AC Q98ND1;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein mlr0194.  
GN MLR0194.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Mesorhizobium.  
OX NCB1\_Taxid=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF030309;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Kawanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
Mesorhizobium loti.";  
RL Mesorhizobium loti.  
DR EMBL; AP002994; BBA47830.1; -  
DR InterPro; IPR002858; DUF124.  
DR ProDom; PD013634; DUF124; 1.

KW Hypothetical protein, Complete proteome.  
 SQ SEQUENCE 221 AA; 24462 MW; 777E5B43B84DF98D CRC64;  
 Query Match 85.4%; Score 35; DB 16; Length 221;  
 Best Local Similarity 71.4%; Pred. No. 26;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVALKRG 9  
 |:|:|:  
 Db 87 PHLKRG 93

RESULT 6  
 ID Q9XF81 PRELIMINARY; PRT; 319 AA.  
 AC Q9XF81;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Aldose reductase ALDRXV4.  
 GN ALDRXV4.  
 OS Xerophyta viscosa.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Velloziaceae; Xerophyta;  
 OC NCBI\_TaxID=30708;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20539414; PubMed=11089682;  
 RA Mundree S.G., Whitaker A., Thomson J.A., Farrant J.M.,  
 RT "An aldose reductase homolog from the resurrection plant Xerophyta  
 viscosa Baker.";  
 RL Planta 211:693-700(2000).  
 DR EMBL: AF133841; AAD2264.1; -  
 DR HSSP: P06632; IHM6.  
 DR InterPro: IPR001395; Aldo/ket red.  
 DR InterPro: IPR000169; SHprot acetyl.  
 DR Pfam: PF00248; Aldo\_ket\_red\_1.  
 DR PRINTS: PR00069; ALDKETREDASE.  
 DR PRODOM: PD000288; Aldo/ket\_red; 1.  
 DR PROSITE: PS00062; ALDO-KETO\_REDUCTASE\_2; 1.  
 DR PROSITE: PS00063; ALDO-KETO\_REDUCTASE\_3; 1.  
 DR PROSITE: PS00639; THIOL\_PROTEASE\_HIS; 1.  
 SQ SEQUENCE 319 AA; 35667 MW; 320761DD85C4B5C CRC64;

Query Match 85.4%; Score 35; DB 10; Length 319;  
 Best Local Similarity 71.4%; Pred. No. 36;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVALKRG 9  
 |:|:|:  
 Db 122 PHLKRG 128

RESULT 7  
 ID Q9WM55 PRELIMINARY; PRT; 855 AA.  
 AC Q9WM55; Q8T9D8;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE CG8013 protein (SD04959p).  
 GN SU(2)12 OR CG8013.  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10711132;  
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Bailew R.M., Baas A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brodeur P., Brodt P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Paulis B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fester C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kemtson J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mileshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusker D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Ceiniker S.E., Adams M.D., Krommiller B., Man K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farrant D.,  
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Idegam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacle J., Paragae V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,  
 RT "Sequencing of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Ceiniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,  
 RT "Annotation of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Flybase:  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [6]

RA SEQUENCE FROM N.A.  
RP Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guatin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nuno J., Pacleb J., Paragas V., Park S., Pounenavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celinker S.,  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AE003515; AAF49094.2; -  
DR EMBL: AY069809; AAL39954.1; -  
DR FLYBase: FBgn0020887; Su(z)12.  
DR InterPro: IPR007087; Znf\_C2H2.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
SQ SEQUENCE 855 AA; 95316 MW; C10FFC6013954105 CRC64;  
  
Query Match 85.4%; Score 35; DB 5; Length 855;  
Best Local Similarity 85.7%; Pred. No. 94;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
OY 3 PVHLKRG 9  
DB 722 PAHLKRG 728  
  
RESULT 8  
O9NJG9 PRELIMINARY; PRT; 900 AA.  
AC O9NJG9:  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Su(z)12.  
GN SU(z)12 OR CG8013.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Birve A., Rasmussen-Lestander A., Larsson J.;  
RT "Suppressor of zeste 12.";  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF149047; AAF73149.1; -  
DR FlyBase: FBgn0020887; Su(z)12.  
DR InterPro: IPR007087; Znf\_C2H2.  
DR SMART: SM00355; Znf\_C2H2\_1;  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
SQ SEQUENCE 900 AA; 100103 MW; 53BA0D83C49EC92F CRC64;  
  
Query Match 85.4%; Score 35; DB 5; Length 900;  
Best Local Similarity 85.7%; Pred. No. 99;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
OY 3 PVHLKRG 9  
DB 722 PAHLKRG 728  
  
RESULT 9  
O9IVC0 PRELIMINARY; PRT; 170 AA.  
AC O9IVC0:  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE RS21-C6-like protein (RS21-C6 protein).  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Spleen.

RA Zhang J., Wang H., Chen W.;  
RL "Unknown protein homolog to murine RS21-C6";  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Zhang J., Chen W.-F.;  
RT "Homology to murine RS21-C6";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AFJ31839; AAK3638.1; -  
DR EMBL: AY029335; AAK37408.1; -  
SQ SEQUENCE 170 AA; 18473 MW; 2D344276C3A1228E CRC64;  
  
Query Match 82.9%; Score 34; DB 11; Length 170;  
Best Local Similarity 85.7%; Pred. No. 32;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
OY 3 PVHLKRG 9  
DB 130 PVHLKRG 136  
  
RESULT 10  
O9QY93 PRELIMINARY; PRT; 170 AA.  
AC O9QY93:  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE RS21-C6 (2410015N17RIK protein) (RIKEN CDNA 2410015N17 gene).  
GN TDRG-TL1 OR 2410015N17RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C;  
RA Wang H., Chen W.-F., Li Y., Jin C.G., Wang Y., Yu Q., Qian X.P.;  
RT "RS21-C6: a novel gene encoding a molecule relevant to TCR and CD3  
expression of pre-T cells.";  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Embryonic stem cells, and Embryo;  
RX MEDLINE=21085660, PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Bateman A., Casavant T.,  
RA Flisemann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuentz P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schirni L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Guetlich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee R.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzaletti J., Mommaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Strauberg R.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF110764; AAF15970.1; -  
DR EMBL: AK010606; BAB27056.1; -  
DR EMBL: AK003643; BAB22909.1; -  
DR EMBL: AK010508; BAB26992.1; -

DR EMBL; BC004623; AA004623.1; -  
DR MGJ; MGJ:1913672; 2410015N17R1K.  
SQ SEQUENCE 170 AA; 18795 MW; A20BCDA7857446A3 CRC64;

Query Match 82.9%; Score 34; DB 11; Length 170;  
Best Local Similarity 85.7%; Pred. No. 32;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 PVHLKRG 9  
DB 130 PVHLKRG 136

## RESULT 11

ID 027878 PRELIMINARY; PRT; 492 AA.

AC 027878.  
DT 01-JUN-1998 (TREMBlrel. 05, Created)  
DT 01-JUN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE Fumarate reductase.  
GN MT1850.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=187420;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=Delta H;  
RX MEDLINE=96037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bashirizadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiwani N., Canuso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
RT delta: functional analysis and comparative genomics.";  
RL J. Bacteriol. 179:7135-7155 (1997).  
CC -1- COFACTOR: BINDS A 2FE-2S CLUSTER (BY SIMILARITY).  
DR EMBL; AE000937; AAB86316.1; -  
DR HSSP; P17596; 10UB.  
DR InterPro; IPR006058; 2FE2S\_ferredoxin.  
DR InterPro; IPR001450; 4FE4S\_ferredoxin.  
DR InterPro; IPR004489; DHB.  
DR InterPro; IPR004017; DUF224.  
DR InterPro; IPR001041; Ferredoxin.  
DR Pfam; PF02754; DUF224; 2.  
DR Pfam; PF00111; fer2; 1.  
DR TIGRFAMs; TIGR00384; dhsb; 1.  
DR TIGRFAMs; TIGR00197; 2FE2S\_FERREDOXIN; 1.  
DR PROSITE; PS00197; 2FE2S\_FERREDOXIN; 1.  
DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 2.  
KW Iron-sulfur; Complete Proteome.

SEQUENCE 492 AA; 54780 MW; 0EF954ED909C5DB CRC64;

Query Match 82.9%; Score 34; DB 17; Length 492;  
Best Local Similarity 85.7%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 PVHLKRG 9  
DB 386 PVHLKRG 392

## RESULT 12

ID 087209 PRELIMINARY; PRT; 492 AA.

AC 087209.  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Succinate dehydrogenase/fumarate reductase Fe-S protein.

GN FRDB/GUPC OR MK0132.  
OS Methanopyrus kandleri.  
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;  
OC Methanopyrus.  
OX NCBI\_TaxID=2320;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
RX MEDLINE=21927647; PubMed=11930014;  
RA Sleatav A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,  
RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
RA Malykh A.G., Koonin E.V., Kozaykin S.A.;  
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
RT and monophyly of archaeal methanogens.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).  
DR EMBL; AE010313; AAM01349.1; -  
DR InterPro; IPR006058; 2FE2S\_ferredoxin.  
DR InterPro; IPR001450; 4FE4S\_ferredoxin.  
DR InterPro; IPR000345; CytC\_heme\_bind.  
DR InterPro; IPR004489; DHB.  
DR InterPro; IPR004017; DUF224.  
DR InterPro; IPR001041; Ferredoxin.  
DR Pfam; PF02754; DUF224; 2.  
DR Pfam; PF00111; fer2; 1.  
DR TIGRFAMs; TIGR00384; dhsb; 1.  
DR PROSITE; PS00197; 2FE2S\_FERREDOXIN; 1.  
DR PROSITE; PS00198; 4FE4S\_FERREDOXIN; 2.  
DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
KW Complete proteome.

SEQUENCE 492 AA; 55720 MW; 4E10FA3C12308618 CRC64;

Query Match 82.9%; Score 34; DB 17; Length 492;  
Best Local Similarity 85.7%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 PVHLKRG 9  
DB 389 PVHLKRG 395

## RESULT 13

ID 099M13 PRELIMINARY; PRT; 86 AA.

AC 099M13.  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Surfactant protein C (Fragment).  
OS Sigmomon hispidus (Hispid cotton rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Sigmodontinae;  
OC Sigmodon.  
OX NCBI\_TaxID=42415;  
RN [1]  
RP SEQUENCE FROM N.A.

RA Bennett J., Guzowski J., Langley R., Pietneva L., Ottolini M.;  
RT "Influenza A Infection Increases Surfactant Protein C Gene Expression  
RT in the Cotton Rat (Sigmomon hispidus).";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.

RA Baco S., Langley R.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF339911; AAK26618.1; -

DR InterPro; IPR001729; Pulm\_surfact\_AP.  
DR ProDom; PD009591; Pulm\_surfact\_AP; 1.  
DR PROSITE; PS00341; SURFACT\_PALMITOYL; 1.  
FT NON\_TER 1  
FT NON\_TER 86

SEQUENCE 86 AA; 9333 MW; E85F946E94639191 CRC64;

Query Match 80.5%; Score 33; DB 11; Length 86;

Best Local Similarity 100.0%; Pred. No. 27;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKR 8  
| | | | |  
DB 12 PVHLKR 17

## RESULT 14

OB1K9  
ID Q81K9 PRELIMINARY; PRT; 231 AA.  
AC Q81K9;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN PF11\_0163.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RX MEDLINE=22255705; PubMed=12368864;  
RA Gardner M.J., Hall N., Fung E., White O., Bertman M., Hyman R.W.,  
RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,  
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
RA Martin D.M.A., Fairhead A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
RA Fraser C.M., Barrrell B.;  
RT "Genome sequence of the human malaria parasite Plasmodium  
falciparum";  
RL Nature 419:498-511(2002).  
DR EMBL; AE014838; AAN35747.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 231 AA; 27908 MW; 5F8414PB9C827554 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 231;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKR 8  
| | | | |  
DB 97 PVHLKR 102

## RESULT 15

Q94F79  
ID Q94F79 PRELIMINARY; PRT; 251 AA.  
AC Q94F79;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Nucleosome/chromatin assembly factor A.  
GN NFA103.  
OS Zea mays (Maize).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. B73;  
RA Chandler V.L., Kaeppler S.M., Kaeppler H.F., Cone K.C.;  
RT "Sequences from the Plant Chromatin Consortium (NSF Plant Genome  
Project 9975930).";  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. B73;

RA Bergstrom D., Springer N.M., Schmitt L.T., Guthrie E., Sidorenko L.,  
RA Selinger D., Kaeppler S.M., Cone K.C.;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF384035; AAK67145.1; -;  
DR InterPro; IPR002164; NAF\_family.  
DR Pfam; PF00956; NAF; 1.  
SQ SEQUENCE 251 AA; 28441 MW; F9E22736CBFA1FC6 CRC64;

Query Match 80.5%; Score 33; DB 10; Length 251;  
Best Local Similarity 85.7%; Pred. No. 75;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 PVHLKRG 9  
| | | | |  
DB 66 PVHLKRG 72

Search completed: December 3, 2003, 15:52:31  
Job time : 8.42857 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using bw model

Run on: December 3, 2003, 15:40:35 ; Search time 73.0884 Seconds  
(without alignments)  
171.565 Million cell updates/sec

Title: US-09-788-308D-2

Perfect score: 413  
Sequence: 1 FPIPLPYCWLKRALIKRIQA.....TLIGRMUPQVCRILVRCSM 79

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	413	100.0	79	22	AAU07651 Human surfactant p
2	413	100.0	286	9	AAAP82933 SAP(Phe), synthe
3	413	100.0	286	9	AAAP80651 Deduced partial se
4	413	100.0	381	8	AAAP70438 Sequence of a huma
5	413	100.0	381	9	AAAP82934 SAP(Phe), synthe
6	413	100.0	381	10	AAAP94756 Sequence of SP-18
7	413	100.0	381	11	AAAP6332 Human alveolar su
8	413	100.0	381	21	AAAY7377 Human surfactant p
9	413	100.0	381	21	AAAY57152 Human surfactant p

10	413	100.0	381	22	AAAB1913
11	413	100.0	382	9	AAAP82982
12	408	98.8	78	9	AAAP80645
13	407	98.5	181	20	AAAB8199
14	407	98.5	243	10	AAAP90600
15	407	98.5	272	21	AAAB34768
16	407	98.5	380	22	AAAB31917
17	407	98.5	381	8	AAAP70664
18	407	98.5	381	11	AAAP04833
19	407	98.5	381	12	AAAB14446
20	407	98.5	381	21	AAAB78944
21	407	98.5	667	22	AAAG2852
22	401	97.1	256	21	AAAY78945
23	401	97.1	256	21	AAAY78945
24	401	97.1	257	11	AAAY78942
25	400	96.9	381	11	AAAB05093
26	398.5	96.5	379	22	AAAB1914
27	395	95.6	293	11	AAAB05418
28	354	85.7	361	7	AAAP60438
29	350	84.7	363	8	AAAP70437
30	347	84.0	362	11	AAAB04211
31	345	83.5	168	21	AAAB34767
32	320	77.5	79	10	AAAP1702
33	316	76.5	60	9	AAAB0584
34	316	76.5	60	12	AAAB10799
35	280	67.8	57	10	AAAP90445
36	271	65.6	61	9	AAAB1076
37	262	63.4	52	9	AAAB0646
38	262	63.4	52	12	AAAB10800
39	213	51.6	40	12	AAAB10798
40	144	34.9	29	12	AAAB10801
41	142	34.4	25	12	AAAB10797
42	142	34.4	25	12	AAAB1623
43	140	33.9	154	21	AAAB58120
44	136	32.9	30	9	AAAB1075
45	130	31.5	26	9	AAAB0644

#### ALIGNMENTS

RESULT 1	AAU07651	AAU07651 standard; Protein: 79 AA.
XX	AAU07651:	
XX	04-DEC-2001 (first entry)	
XX	Human surfactant protein B (SP-B).	
XX	Human surfactant protein B; surfactant protein C; SP-B; SP-C; peptoid;	
XX	pulmonary; protein therapy; spreading agent; N-substituted glycine;	
XX	lung surfactant; pulmonary surfactant; alveolar surface activity;	
XX	respiratory distress syndrome.	
XX	Homo sapiens.	
XX	MO200160837-A2.	
XX	23-AUG-2001:	
XX	16-FEB-2001; 2001MO-US05145.	
XX	16-FEB-2001; 2000US-0182847.	
XX	(NOUN) UNIV NORTHWESTERN.	
XX	(CHIR) CHIRON CORP.	
XX	PI, Barjon AE, Zuckerman RN, Wu CW;	
XX	WPI; 2001-550045/61.	
XX		

Amino acid sequenc  
Human SP18 deduced  
Synthetic hydropho  
Human SP-18 surfac  
Sequence of human  
Human secreted pro  
Amino acid sequenc  
6kd pulmonary surf  
Includes a 6kd hum  
Pulmonary surfacta  
Human SP-B prepro  
Novel human diagno  
Mature alveolar su  
Synthetic alveolar  
Alveolar surfactan  
Gene product of ve  
Amino acid sequenc  
CAT-SP-B hybrid pr  
Dog 10 kd alveolar  
Sequence of a cani  
Deduced amino acid  
Human secreted pro  
Protein increasing  
Human SAP (Phe) pep  
Pulmonary surfacta  
Human pulmonary su  
Sequence encoded b  
Synthetic hydropho  
Pulmonary surfacta  
Pulmonary surfacta  
Pulmonary surfacta  
Pulmonary surfacta  
N-terminal 25 amin  
Lung cancer associ  
Sequence encoded b  
Synthetic hydropho

PT Heteropolymeric pulmonary spreading agent having at least one  
 PT N-substituted glycine residue and an amino acid residue corresponding  
 PT to a natural surfactant-associated protein, useful for treating lung  
 PT respiratory distress -  
 XX  
 XX Claim 1: Fig 3; 40pp; English.  
 XX  
 CC The invention relates to a non-natural heteropolymeric pulmonary  
 CC spreading agent comprising at least one N-substituted glycine residue and  
 CC at least one amino acid residue corresponding to the surfactant proteins  
 CC B and C (SP-B and SP-C). These surfactant protein mimics are peptid  
 CC sequences (reverse sequence of the natural protein i.e. equal to the  
 CC carboxy to amino sequence of the peptide) added to a lipid mixture to  
 CC create a functional, non-immunogenic lung surfactant with physiological  
 CC alveolar surface activity. The peptid sequences of the invention can be  
 CC used to enhance the solubility of surfactant associated proteins (to  
 CC therefore enhance resistance to aggregation) and can also affect alveolar  
 CC surface tension during an inhalation/exhalation cycle. The spreading  
 CC agents are useful for treating disorders of the lungs such as respiratory  
 CC distress syndrome. This sequence represents the human surfactant protein  
 CC B (SP-B).  
 CC  
 XX  
 XX Sequence 79 AA;  
 SQ  
 Query Match 100.0%; Score 413; DB 22; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-44;  
 Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FPIPIPYCMLCRALIKRIQAMIPKGLRVAVAQVCRVPLVAGGICQCLAERYSVILLDT 60  
 DB 1 FPIPIPYCMLCRALIKRIQAMIPKGLRVAVAQVCRVPLVAGGICQCLAERYSVILLDT 60  
 QY 61 LLGRMLPOLVCRVLVRCSM 79  
 DB 61 LLGRMLPOLVCRVLVRCSM 79  
 RESULT 2  
 AAP82933  
 ID AAP82933 standard; protein; 286 AA.  
 XX  
 AC AAP82933;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 21-NOV-1990 (first entry)  
 XX  
 DE SAP (Phe).  
 DE  
 XX SAP (Val); SAP (Phe); hyaline membrane disease.  
 KM  
 XX  
 OS synthetic.  
 OS  
 XX WO8804324-A.  
 PN  
 XX 16-JUN-1988;  
 PD  
 XX 03-DEC-1987; 87WO-US03180.  
 PF  
 XX 08-DEC-1986; 86US-0939206.  
 PR 10-JUN-1987; 87US-0060719.  
 PR 01-OCT-1987; 87US-0101680.  
 XX  
 XX (ABBO ) ABBOTT LABORATORIES.  
 PA  
 XX  
 PI Whitsett JA, Fox JL, Pilotmatia TJ, Meuth JL;  
 DR WPI; 1988-175472/25.  
 DR N-PSDB; AAN80613.  
 PT Pulmonary hydrophobic surfactant-associated proteins - useful for  
 PT normalising pulmonary surface tension.  
 XX  
 PS Disclosure; Page 7; 7pp; English.

XX This pulmonary hydrophobic surfactant-associated protein SAP(Phe)  
 CC is encoded by a partial cDNA clone. When SAP (Phe) is combined with  
 CC a lipid it can be used to reduce or maintain normal pulmonary surface  
 CC tension in the alveoli of animals (esp. humans) and therefore can be  
 CC used in the treatment of hyaline membrane disease in premature infants.  
 CC It may also be used to deliver substances to respiratory epithelial  
 CC cells. See also AAP80572-96, AAN80615-17 and AAN82412-13.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 CC (Updated on 25-MAR-2003 to correct PI field.)  
 CC  
 XX  
 XX Sequence 286 AA;  
 SQ  
 Query Match 100.0%; Score 413; DB 9; Length 286;  
 Best Local Similarity 100.0%; Pred. No. 5.6e-44;  
 Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 FPIPIPYCMLCRALIKRIQAMIPKGLRVAVAQVCRVPLVAGGICQCLAERYSVILLDT 60  
 DB 106 FPIPIPYCMLCRALIKRIQAMIPKGLRVAVAQVCRVPLVAGGICQCLAERYSVILLDT 165  
 QY 61 LLGRMLPOLVCRVLVRCSM 79  
 DB 166 LLGRMLPOLVCRVLVRCSM 184  
 RESULT 3  
 AAP80651  
 ID AAP80651 standard; protein; 286 AA.  
 XX  
 AC AAP80651;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 17-SEP-1990 (first entry)  
 XX  
 DE Deduced partial sequence of human pulmonary hydrophobic surfactant-  
 DE associated protein (SAP) (Phe).  
 DE  
 XX Human pulmonary hydrophobic surfactant-associated protein (SAP) (Phe);  
 KM hyaline membrane disease (HMD) prevention; assays.  
 KM  
 XX Homo sapiens.  
 OS  
 XX  
 FT Key Location/Qualifiers  
 FT Region 106..119  
 FT /note="corresponds to N-terminal of human SAP (Phe) "  
 FT Region 134..152  
 FT /note="hydrophobic region"  
 XX  
 PN WO8803170-A.  
 PN  
 PD 05-MAY-1988.  
 PF  
 XX 02-OCT-1987; 87WO-US02536.  
 PF  
 XX 08-DEC-1986; 86US-0939206.  
 PR 10-JUN-1987; 87US-0060719.  
 PR 01-OCT-1987; 87US-0101680.  
 XX  
 XX (WHITT/) WHITSETT J A.  
 PA (ABBO ) ABBOTT LAB.  
 PA  
 XX  
 PI Whitsett JA, Fox JL, Pilotmatia TJ, Meuth JL, Sarin VK;  
 DR WPI; 1988-133244/19.  
 DR N-PSDB; AAN80640.  
 PT Pulmonary hydrophobic surfactant-associated proteins -  
 PT used with lipid(s) to treat and prevent hyaline membrane disease  
 PT and similar syndromes  
 XX  
 PS Example; Fig 3a-3b; 144pp; English.



CC It is deduced from the sequence of a partial cDNA clone. The hydrophobic  
CC regions of SAP (Phe) (AAP80651) and SAP (Val) (AAP80652) are somewhat  
CC homologous. Although these two proteins are encoded by distinct genes,  
CC it is believed that they are structurally related. SAP (Val) and SAP  
CC (Phe), when combined with lipids, have significant pulmonary biophysical  
CC surfactant activity that may be utilized to treat and prevent hyaline  
CC membrane disease (HMD) and other syndromes associated with lack of  
CC insufficient amcs. of natural pulmonary surfactant material. Antibodies  
CC and antiserum may also be made which are directed against SAP (Val) or  
CC SAP (Phe). SAP (Val) and SAP (Phe) in body fluids may be assayed using  
CC the compans.  
CC (Updated on 25-MAR-2003 to correct PR field.)  
CC (Updated on 25-MAR-2003 to correct PA field.)  
CC (Updated on 25-MAR-2003 to correct PI field.)  
XX

Query Match	100.0%	Score 413	DB 9	Length 286
Best Local Similarity	100.0%	Pred. 5.6e-44		
Matches 79	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY 1 FPIPLPYCMLCRALIKRIQAMIPKGLRVAAQVCRVPLVAGGICQCLAERYSVILTD 60  
|||  
Db 106 FPIPLPYCMLCRALIKRIQAMIPKGLRVAAQVCRVPLVAGGICQCLAERYSVILTD 165

QY 61 LIGRLPQLVCRLVLRCSM 79  
|||  
Db 166 LIGRLPQLVCRLVLRCSM 184

RESULT 4  
AAP70438  
ID AAP70438 standard; protein; 381 AA

XX	25-MAR-2003	(updated)
DT	17-JAN-1991	(first entry)
DT		

DE Sequence of a human 18 kd alveolar surfactant protein (ASP, from clone #3.

**KW** Lung surfactant; respiratory disease syndrome; therapy

OS Homo sapiens

PN W08706588-A.

PD 05-NOV-1987

PF 30-APR-1987; 87WO-US00978

PR 30-APR-1986; 86US-0857715

XX  
XX /CARD ) CARTRODUTA PROMOZIONE DAXX  
PI Cob:11:as TM

PI White RT;  
PT Cordell P.

PI Benson BJ;  
yy

DR WPI; 1987-320974/45  
DR N-PSDB: A8N70697

XX	Pure alveolar surface
PT	

PT and affinity chromatography for treating respiratory disease

XX	Claim 2. Filed 2. 73nn. English
PS	

XX An ASP encoded by human SP-18 DNA and human SP-5 DNA is claimed

The purified ASP is suitable for treating respiratory diseases

CC	syndrome in mammals, esp. when administered with a phospholipid
CC	and opt. with the 33X ASP protein. Human and canine ASP-18 differ
CC	only slightly in AA sequence.
CC	(updated on 25-MAR-2003 to correct PR field.)
CC	(updated on 25-MAR-2003 to correct PA field.)
XX	Sequence 381 AA;

Query Match	100.0%	Score 413	DB 8	Length 381
Best Local Similarity	100.0%	Pred. No. 7	8e-44	
Matches	79	Conservative	0	Mismatches 0; Indels 0; Gaps 0

**Dy**      1 FPIPLPYCMLCRALIKRIQAMIPKCALRVAQAQVNVPLVAGGICQCCLAERYSVILDDT 60  
         |||||  
**Db**      201 FPIPLPYCMLCRALIKRIQAMIPKCALRVAQAQVNVPLVAGGICQCCLAERYSVILDDT 260

QY	61	LGRLPQLVCRVLRC	SM	79
Db	261	LGRLPQLVCRVLRC	SM	279

RESULT 5	
AAP82934	
ID AAP82934	standard; protein; 381 AA

XX	25-MAR-2003	(updated)
DT	21-NOV-1990	(first entry)
DT		

DE	SAP (Phe) .
XX	
KW	SAP (Val) ; SAP (Phe) ; hyaline membrane disease

OS           synthetic.  
XX  
PN           W08804324-A

PD	16-JUN-1988
XX	
PF	03-DEC-1987

AA		
PR	08-DEC-1986;	86US-0939206
PR	10-JUN-1987;	87US-0060719
PR	01-OCT-1987;	87US-0101680

XX (ABBO ) ABBOTT LABORATORIES.  
PA  
XX  
PI Whiteett JA Fox JJ Pilot

XX  
DR  
DR  
WPI; 1988-175472/25.  
N-PSDB; AAN80615, AAN80616.

Pulmonary hydrophobic surfactant-associated proteins - useful for normalising pulmonary surface tension.

This pulmonary hydrophobic surfactant-associated protein SAP(Phe) is encoded by a full-length cDNA clone. When SAP(Phe) is combined with

CC a lipid it can be used to reduce or maintain normal pulmonary surface  
CC tension in the alveoli of animals (esp. humans) and therefore can be  
CC used in the treatment of hyaline membrane disease in premature infants  
CC It may also be used to deliver substances to neonatal airways

CC cells. See also AAP80572-96, AAN80615-17 and AAN82412-13.  
CC (Updated on 25-MAR-2003 to correct PR field.)  
CC (Updated on 25-MAR-2003 to correct PI field.)

**SQ** Sequence 381 AA;

Query Match	Score	DB	Length
100.0%	413	9	381

Best Local Similarity	100.0%;	Pred. No. 7.8e-44;
Matches 79;	Conservative 0;	Mismatches 0;
Indels 0;	Gaps 0;	

QY 1 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAVQVCRVPLVAGGICQCLAERYSVILLDT 60  
DB 201 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAVQVCRVPLVAGGICQCLAERYSVILLDT 260  
QY 61 LLGRMLPOLVCRVLVLRCSM 79  
DB 261 LLGRMLPOLVCRVLVLRCSM 279

RESULT 6  
AAP94756  
ID AAP94756 standard; protein; 381 AA.

AC AAP94756;  
XX  
XX 25-MAR-2003 (updated)  
DT 19-JAN-1991 (first entry)  
XX  
DE Sequence of SP-18 precursor protein encoded by human CDNA#3.  
XX  
XX Alveolar surfactant protein; respiratory distress syndrome;  
XX pneumonia; bronchitis; therapy.  
XX  
XX Homo sapiens.

XX  
XX Key Location/Qualifiers  
FT Peptide 1..200  
PT Protein 201..381

XX W08904326-A.  
XX  
XX 18-MAY-1989.

XX 02-NOV-1988; 88WO-US03899.

XX 04-NOV-1987; 87US-0117099.  
XX 01-NOV-1988; 88US-0266443.

XX (CALD) CALIFORNIA BIOTECHNOLOGY INC.

XX Benson BJ, White RT, Schilling JW, Buckley D, Scarborough RM,  
XX  
XX WPI; 1989-165617/22.  
XX N-PSDB; AAN93635.

XX Human SP-18 and SP-5 derived peptide(s) -  
PT with alveolar surfactant protein activity, used for treating  
PT respiratory distress syndrome, pneumonia and bronchitis  
XX

XX Disclosure; Fig 4-1 -4-2; .PP; English.

XX ASP proteins including the hsp-18- and hsp-5-derived peptides can be  
CC used as a carrier or vehicle for delivery of other active and important  
CC molecules to and/or through the lung to the blood vasculature.  
CC (updated on 25-MAR-2003 to correct PR field.)  
CC (updated on 25-MAR-2003 to correct PA field.)  
XX

XX Sequence 381 AA.

QY Query Match 100.0%; Score 413; DB 10; Length 381;  
Best Local Similarity 100.0%; Pred. No. 7.8e-44;  
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAVQVCRVPLVAGGICQCLAERYSVILLDT 60  
DB 201 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAVQVCRVPLVAGGICQCLAERYSVILLDT 260

QY 61 LLGRMLPOLVCRVLVLRCSM 79  
DB 261 LLGRMLPOLVCRVLVLRCSM 279

RESULT 7  
AAR06332  
ID AAR06332 standard; protein; 381 AA.

AC AAR06332;  
XX  
XX 07-DEC-1990 (first entry)  
DT  
XX

XX Human alveolar surfactant protein (SP-B).

XX Alveolar surfactant; SP-A; SP-B; SP-C; respiratory distress syndrome;  
XX oxygen toxicity; alpha-1-anti-protease; emphysema; lung cancer;  
XX bronchitis; asthma; tuberculosis;  
XX

XX Homo sapiens.

XX W09007469-A.  
XX  
XX 12-JUL-1990.

XX 29-DEC-1989; 89WO-US00587.

XX 29-DEC-1988; 88US-0295926.

XX (BENS/) BENSON B J.

XX Benson BJ, Wright J;  
XX  
XX WPI; 1990-238980/31.

XX Pulmonary admin. of liposome contg. active cpds. - uses alveolar  
PT surfactant protein to enhance transport across lung surface  
PT useful for treating variety of lung specific diseases  
XX

XX Disclosure; Page 7; ?pp; English.  
XX  
XX Alveolar surfactant protein is useful in enhancing the uptake of  
CC liposomes containing a pharmaceutically active compound (local or  
CC systemic) across the pulmonary surface.  
CC This is useful in treating a variety of lung specific diseases  
CC eg. respiratory distress syndromes, pneumonia, oxygen toxicity,  
CC alpha-1-anti-protease deficiency, emphysema, asthma, tuberculosis,  
CC lung cancer and bronchitis.  
XX

XX Sequence 381 AA;

QY Query Match 100.0%; Score 413; DB 11; Length 381;  
Best Local Similarity 100.0%; Pred. No. 7.8e-44;  
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAVQVCRVPLVAGGICQCLAERYSVILLDT 60  
DB 201 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAVQVCRVPLVAGGICQCLAERYSVILLDT 260

QY 61 LLGRMLPOLVCRVLVLRCSM 79  
DB 261 LLGRMLPOLVCRVLVLRCSM 279

RESULT 8  
AA97377

ID AA97377 standard; protein; 381 AA.

AC AA97377;  
XX  
XX 14-SEP-2000 (first entry)  
DT  
XX

XX Human surfactant protein-B.

XX Human; bacterial infection; inflammation; cystic fibrosis;  
XX gastrointestinal tract infection; pulmonary surfactant; SP-B;  
XX surfactant protein-B; antibacterial; fusion protein; lysosome.  
XX

OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..23  
 FT Peptide /label= signal\_peptide  
 FT Peptide 24..200  
 FT Protein /label= propeptide  
 FT Protein 201..279  
 FT Peptide /label= mature\_SP-B  
 FT Peptide 280..381  
 FT Peptide /label= propeptide  
 XX  
 PN MO200029588-A1.  
 XX  
 PD 25-MAY-2000.  
 XX  
 PF 18-NOV-1999; 99MO-US27403.  
 XX  
 PR 18-NOV-1998; 98US-0193877.  
 PR 16-NOV-1999; 99US-0440742.  
 XX  
 PA (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.  
 XX  
 PI Weaver TE, Akindi HT;  
 XX  
 DR WPI; 2000-387800/33.  
 XX  
 PT Lysozyme/surfactant protein-B fusion proteins for prophylaxis and  
 PT treatment of bacterial infections, respiratory infections such as  
 PT cystic fibrosis and gastro intestinal infections in a mammal -  
 XX  
 PS Disclosure; Page 33-34; 42pp; English.  
 XX  
 CC The present sequence is human surfactant protein-B (SP-B). It was  
 CC used to create a fusion protein with recombinant rat lysozyme. Lysozyme  
 CC acts in an antibacterial manner in vitro and, in doing so, reduces the  
 CC inflammation associated with infection. SP-B is necessary for normal  
 CC respiratory function. The fusion protein can be used to treat bacterial  
 CC infection in the respiratory and gastrointestinal tracts. In particular,  
 CC it can be used to treat infection by Pseudomonas aeruginosa in cystic  
 CC fibrosis sufferers.  
 CC  
 SQ Sequence 381 AA;  
 XX  
 QY Query Match 100.0%; Score 413; DB 21; Length 381;  
 DB Best Local Similarity 100.0%; Pred. No. 7, 8e-44;  
 QY Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 1 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAQVCRVPLVAGGICQCLAEKRSVILLDT 60  
 201 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAQVCRVPLVAGGICQCLAEKRSVILLDT 260  
 QY 61 LIGRMPLQVLCRLVLRCSM 79  
 DB 261 LIGRMPLQVLCRLVLRCSM 279  
 XX  
 RESULT 9  
 AAY57152  
 ID AAY57152 standard; protein, 381 AA.  
 XX  
 AC AAY57152;  
 XX  
 DT 11-FEB-2000 (first entry)  
 XX  
 DE Human surfactant protein-B (SP-B) sequence.  
 XX  
 KW Lysozyme/surfactant protein-B fusion protein; bacterial infection;  
 KW lysozyme; chemotaxis; inflammation; antibacterial; gastrointestinal;  
 KW respiratory; cystic fibrosis; surfactant protein-B; SP-B.  
 XX  
 OS Homo sapiens.  
 XX

PN US5993809-A.  
 XX  
 PD 30-NOV-1999.  
 XX  
 PF 18-NOV-1998; 98US-0193877.  
 XX  
 PR 18-NOV-1998; 98US-0193877.  
 XX  
 PA (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.  
 XX  
 PI Weaver TE, Akindi HT;  
 XX  
 DR WPI; 2000-038237/03.  
 XX  
 PT Lysozyme/surfactant protein-B fusion proteins useful for treating  
 PT bacterial infections of the respiratory system -  
 XX  
 PS Disclosure; Columns 7-10; 10pp; English.  
 XX  
 CC The invention provides a composition comprising a lysozyme/surfactant  
 CC protein-B fusion protein (sequences AAY57153 or AAY57156) that can be  
 CC used in a method of prophylaxis or treatment of a bacterial infection in  
 CC a mammal. The lysozyme component of the fusion protein is a naturally  
 CC occurring antibacterial agent which kills bacteria by hydrolyzing the  
 CC glycosidic bond between C-1 of N-acetylmuramic acid and C-4 of the  
 CC N-acetylglucosamine in the bacterial polysaccharide cell wall. It may  
 CC also act synergistically with complement factors and antibodies to lyse  
 CC bacterial cells. Lysozyme also exhibits chemotaxis of polymorphonuclear  
 CC leukocytes and limits the production of oxygen free radicals following  
 CC an infection. This limits the degree of inflammation and enhances  
 CC phagocytosis by the cells. Lysozyme is probably also implicated in the  
 CC response of airway tissue to injury. The composition may be used for the  
 CC prophylaxis or treatment of a bacterial infection (especially  
 CC respiratory bacterial infection (e.g. cystic fibrosis) or  
 CC gastrointestinal bacterial infection) in a mammal. The method and the  
 CC composition eliminate problems associated with conventional antibiotic  
 CC treatments such as inefficacy and promotion of antibiotic resistant  
 CC bacterial strains. The present sequence represents a human surfactant  
 CC protein-B (SP-B) used in the construction of the fusion protein of the  
 CC invention. SP-B is synthesized as a propeptide by the alveolar type II  
 CC epithelial cells.  
 CC  
 SQ Sequence 381 AA;  
 XX  
 QY Query Match 100.0%; Score 413; DB 21; Length 381;  
 DB Best Local Similarity 100.0%; Pred. No. 7, 8e-44;  
 QY Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 1 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAQVCRVPLVAGGICQCLAEKRSVILLDT 60  
 201 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAQVCRVPLVAGGICQCLAEKRSVILLDT 260  
 QY 61 LIGRMPLQVLCRLVLRCSM 79  
 DB 261 LIGRMPLQVLCRLVLRCSM 279  
 XX  
 RESULT 10  
 AAB31913  
 ID AAB31913 standard; protein, 381 AA.  
 XX  
 AC AAB31913;  
 XX  
 DT 15-MAY-2001 (first entry)  
 XX  
 DE Amino acid sequence of a human protein.  
 XX  
 KW Human; perlecan; retinol-binding plasma protein; calgranulin B; vaccine;  
 KW ganglioside GM2 activator; saposin B; degenerative disease; glial cell;  
 KW neurological disease; auto-immune disease; multiple sclerosis; toxicity;  
 KW Alzheimer's disease; Parkinson's disease; amyotrophic lateral sclerosis;  
 KW rheumatoid polyarthritis; lupus erythematosus; gene therapy.  
 XX

```
OS Homo sapiens.
XX WO200105422-A2.
XX 25-JAN-2001.
XX 17-JUL-2000; 2000WO-FR02057.
XX 15-JUL-1999; 99FR-0009372.
XX (INMR ) BIOMERIEUX STELHYS.
XX Roecklin D, Kolbe H, Charles M, Malcus C, Santoro L, Perron H;
XX WPI; 2001-159475/16.
XX
XX Detecting, preventing and treating degenerative, neurological and
XX autoimmune diseases, particularly multiple sclerosis, using specified
XX polypeptides or related nucleic acid or ligand
XX
XX Claim 1; Page 169-170; 209pp; French.
XX
XX The present sequence represents a human protein, which is used in the
XX method of the invention. The specification describes a method which uses
XX at least one polypeptide or polynucleotide sequence belonging to the
XX perlecan, precursor of the retinol-binding plasma protein, precursor of
XX the ganglioside GM2 activator, calgranulin B or saposin B protein
XX families. The method is used for detecting, preventing or treating a
XX degenerative, neurological and/or auto-immune disease. The
XX polynucleotides and polypeptides are used for diagnosis, prognosis,
XX prevention and treatment of multiple sclerosis (in its various forms
XX and phases). They may also be useful in cases of e.g. Alzheimer's and
XX Parkinson's diseases, amyotrophic lateral sclerosis, rheumatoid
XX polyarthritis and lupus erythematosus, including use as vaccines and
XX in gene therapy (expression of sense or antisense sequences). They can
XX also be used to assess efficacy of potential therapeutic agents. They can
XX particularly compounds that reduce or inhibit toxicity towards glial
XX cells.
XX
XX Sequence 381 AA;
XX
XX Query Match 100.0%; Score 413; DB 22; Length 381;
XX Best Local Similarity 100.0%; Pred. No. 7.8e-44;
XX Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 FPIPLPYCWLCRALIKRIQAMIPKGLRVAAVAVQCRVPLVAGGICCCCLARYSVILLDT 60
XX DB 201 FPIPLPYCWLCRALIKRIQAMIPKGLRVAAVAVQCRVPLVAGGICCCCLARYSVILLDT 260
XX
XX 61 LIGRMLPOLVCRVLVRCSCM 79
XX DB 261 LIGRMLPOLVCRVLVRCSCM 279
XX
XX
XX RESULT 11
XX AAP82982 standard; protein; 382 AA.
XX AAP82982;
XX
XX 25-MAR-2003 (updated)
XX 04-DEC-1990 (first entry)
XX
XX Human SP18 deduced from cDNA clone #3.
XX
XX Alveolar surfactant protein; ASP; respiratory distress syndrome;
XX pneumonia; human; bronchitis; SP18.
XX
XX Canus familiaris.
XX
XX Key Location/Qualifiers
XX FT Protein 201..286
XX /label=mature processed protein
```

```
FT Misc-difference 131
FT /label=T or I
FT /note="T in genomic clone; I in cDNA clone"
FT Modified-site 129..131
FT /label=N-glycosylation site
FT note="only in genomic clone"
FT 311..313
FT /label=N-glycosylation site
XX
XX WO8805820-A.
XX
XX 11-AUG-1988.
XX
XX 15-JAN-1987; 87WO-US00092.
XX
XX 30-APR-1986; 86US-0857715.
XX 29-JAN-1987; 87US-0008453.
XX
XX (CALB-) CALIF BIOTECHN INC.
XX
XX Schilling JW, White RT, Cordell B, Benson BJ;
XX WPI; 1988-124493/33.
XX
XX N-PSDB; AAN80703.
XX
XX Recombinant alveolar surfactant protein - used for treating
XX respiratory distress syndrome and related diseases e.g.
XX pneumonia and bronchitis.
XX
XX Disclosure; Page ?; ?pp; English.
XX
XX The sequence was deduced from the sequence of a clone isolated
XX from a human genomic cDNA library. The protein is the alveolar
XX surfactant protein SP18, a member of the low mol. wt., hydro-
XX phobic 10K ASP group. The genomic and cDNA sequences encode
XX sequences which differ by a single residue, Ile131 of the cDNA
XX is Thr in the genomic clone. Thus the genomic clone-encoded
XX precursor contains 2 N-glycosylation sites and the cDNA-encoded
XX recombinant protein and used for the treatment of respiratory
XX disorders.
XX See also AAP82977-80 and AAP80694.
XX (Updated on 25-MAR-2003 to correct PR field.)
XX (Updated on 25-MAR-2003 to correct DR field.)
XX
XX Sequence 382 AA;
XX
XX Query Match 100.0%; Score 413; DB 9; Length 382;
XX Best Local Similarity 100.0%; Pred. No. 7.8e-44;
XX Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 FPIPLPYCWLCRALIKRIQAMIPKGLRVAAVAVQCRVPLVAGGICCCCLARYSVILLDT 60
XX DB 202 FPIPLPYCWLCRALIKRIQAMIPKGLRVAAVAVQCRVPLVAGGICCCCLARYSVILLDT 261
XX
XX 61 LIGRMLPOLVCRVLVRCSCM 79
XX DB 262 LIGRMLPOLVCRVLVRCSCM 280
XX
XX
XX RESULT 12
XX AAP80645 standard; protein; 78 AA.
XX AAP80645;
XX
XX 25-MAR-2003 (updated)
XX 17-SEP-1990 (first entry)
XX
XX Synthetic hydrophobic surfactant-associated protein (SAP) (Phe) peptide.
XX Synthetic hydrophobic surfactant-associated protein (SAP) (Phe);
XX hyaline membrane disease (HMD) prevention; assays.
XX
```

XX OS Homo sapiens.  
XX PN MO8803170-A.  
XX PD 05-MAY-1988.  
XX PF 02-OCT-1987; 87WO-US02536.  
XX PR 08-DEC-1986; 86US-0939206.  
XX PR 10-JUN-1987; 87US-0060719.  
XX PR 01-OCT-1987; 87US-0101680.  
XX PA (WHIT/) WHITSETT J A.  
XX PA (ABBO ) ABBOTT LAB.  
XX PI Whiteett JA, Fox JL, Pilotmatia TJ, Meuth JL, Sarin VK;  
XX DR WPI; 1988-133244/19.  
XX PT Pulmonary hydrophobic surfactant-associated proteins -  
XX PT used with lipid(s) to treat and prevent hyaline membrane disease  
XX and similar syndromes  
XX  
XX PS Claim 27, Page 86; 144pp; English.  
XX  
XX CC It can be made by chemical or enzymatic peptide synthesis. Also claimed  
XX CC is a purified and isolated DNA sequence encoding SAP (Val). SAP (Val)  
XX CC and SAP (Phe), when combined with lipid, have significant pulmonary  
XX CC biophysical surfactant activity that may be used to treat and prevent  
XX CC hyaline membrane disease (HMD) and other syndromes associated with lack  
XX CC of insufficient amts. of natural pulmonary surfactant material.  
XX CC Antipodols and antisera may also be made which are directed against SAP  
XX CC (Val) or SAP (Phe). SAP (Val) and SAP (Phe) in body fluids may be  
XX CC assayed using the compens.  
XX CC (Updated on 25-MAR-2003 to correct PR field.)  
XX CC (Updated on 25-MAR-2003 to correct PA field.)  
XX CC (Updated on 25-MAR-2003 to correct PI field.)  
XX SQ Sequence 78 AA;  
  
Query Match 98.8%; Score 408; DB 9; Length 78;  
Best Local Similarity 100.0%; Pred. NO. 5.6e-44;  
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAVQVCRVPLVAGGICQCLAERYSVILLDT 60  
DB 1 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAVQVCRVPLVAGGICQCLAERYSVILLDT 60  
QY 61 LIGRMPLPOLVCRVLVRCSS 78  
DB 61 LIGRMPLPOLVCRVLVRCSS 78  
  
RESULT 13  
AAW88199  
ID AAW88199 standard; Protein; 181 AA.  
XX  
AC AAW88199;  
XX  
DT 15-MAR-1999 (first entry)  
XX  
DE Human SP-18 surfactant protein.  
XX  
KM Surfactant; SP-18; SP-B; pulmonary lavage; inflammation;  
KM acute hyposmia; congenital diaphragmatic hernia;  
KM respiratory distress syndrome; meconium aspiration syndrome;  
KM pneumonia; therapy.  
XX  
OS Homo sapiens.  
XX MO9849191-A1.  
XX

PD 05-NOV-1998.  
XX  
XX PF 29-JAN-1998; 98WO-US01711.  
XX PR 28-APR-1997; 97US-0848580.  
XX PA (SCRI ) SCRIPPS RES INST.  
XX PI Cochrane CG, Revak SD;  
XX DR WPI; 1999-034654/03.  
XX DR N-PSDB; AAV65196.  
XX  
XX PT Pulmonary lavage with dilute surfactant solution at positive  
XX PT end-expiratory pressure - with removal of fluid using short periods  
XX PT of suction, used to treat respiratory distress, e.g. in neonates  
XX PT where caused by aspiration of meconium  
XX  
XX PS Disclosure; Page 117; 145pp; English.  
XX  
XX CC This is human SP-18 (SP-B) surfactant protein. A claimed method  
XX CC for pulmonary lavage comprises applying gas positive and  
XX CC end-expiratory pressure into a lung section of a mammal at a  
XX CC pressure of 4-16 cm of water, instilling a lavage composition  
XX CC containing dilute surfactant in aqueous medium to the lung, and  
XX CC removing pulmonary fluid from the lung using short intervals of  
XX CC tracheo-bronchial suction at a negative pressure of about 20-100 mm  
XX CC of mercury. The lavage composition may contain a natural pulmonary  
XX CC surfactant such as SP-B or SP-C, or a synthetic surfactant composed  
XX CC of phospholipids and synthetic peptides (see AAW82278-97). Lavage is  
XX CC used to treat respiratory distress syndrome caused by aspiration of  
XX CC meconium or gastric contents, pulmonary inflammation or infection,  
XX CC acute hypoxemia, persistent foetal circulation, congenital  
XX CC diaphragmatic hernia, sepsis, trauma, pancreatitis, inhalation of  
XX CC hot or noxious vapour, pneumonia or multiple transfusions. The  
XX CC lavage solution removes inflammatory mediators and preserves or  
XX CC restores pulmonary function.  
XX  
XX SQ Sequence 181 AA;  
  
Query Match 98.5%; Score 407; DB 20; Length 181;  
Best Local Similarity 98.7%; Pred. NO. 1.9e-43;  
Matches 78; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAVQVCRVPLVAGGICQCLAERYSVILLDT 60  
DB 1 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAVQVCRVPLVAGGICQCLAERYSVILLDT 60  
QY 61 LIGRMPLPOLVCRVLVRCSSM 79  
DB 61 LIGRMPLPOLVCRVLVRCSSM 79  
  
RESULT 14  
AAP90600  
ID AAP90600 standard; Protein; 243 AA.  
XX  
AC AAP90600;  
XX  
DT 25-MAR-2003 (updated)  
DT 31-OCT-1989 (first entry)  
XX  
DE Sequence of human SP18 monomer.  
XX  
KM Human SP18 monomer; respiratory distress syndrome.  
KM Homo sapiens.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FH Peptide 63..143  
FT Peptide 63..77  
FT Peptide 73..87  
FT Peptide 83..97

FT Peptide 93..107  
 FT Peptide 103..117  
 FT Peptide 108..138  
 FT Peptide 113..127  
 FT Peptide 113..134  
 FT Peptide 113..138  
 FT Peptide 116..134  
 FT Peptide 116..138  
 FT Peptide 123..137  
 FT Peptide 133..143  
 FT Peptide 128..143  
 FT Peptide 121..143  
 FT Peptide 114..143  
 FT Peptide 113..143  
 FT Peptide 113..142  
 FT Peptide 98..143  
 FT Peptide 94..143  
 PN W08906657-A.  
 PD 27-JUL-1989.  
 XX 05-JAN-1989; 89WO-US00046.  
 XX 06-JAN-1988; 88US-0141200.  
 PR 04-JAN-1989; 89US-0293201.  
 XX (SCRI) SCRIPPS CLINIC & RES FOUND.  
 PA Cochrane CG, Revak SD;  
 PI WPI; 1989-233825/32.  
 DR N-PSDB; AAN90417.  
 XX Human SPI8 monomer protein and gene - used for preparing synthetic  
 PT pulmonary surfactants for treatment of respiratory distress syndrome.  
 XX Claim 1; page 70; and fig 1; 89pp; English.  
 XX The protein (mature protein is residues 187-429) and fragments of it  
 CC (see features) may be mixed with a phospholipid to form a synthetic  
 CC pulmonary surfactant, which can be used to treat respiratory distress  
 CC syndrome. See AAN90417, and AAP90605-11.  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 CC  
 XX Sequence 243 AA;  
 SQ  
 Query Match 98.5%; Score 407; DB 10; Length 243;  
 Best Local Similarity 98.7%; Pred. No. 2.7e-43;  
 Matches 78; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 1 FPIPLPYCMLCRALIKRIQAMIPKGLRVAVAVQVGVPLVAGGICQCLAERYVILLDT 60  
 Db 63 FPIPLPYCMLCRALIKRIQAMIPKGLRVAVAVQVGVPLVAGGICQCLAERYVILLDT 122  
 Oy 61 LIGRMLPOLVCRVLRCSM 79  
 Db 123 LIGRMLPOLVCRVLRCSM 141  
 RESULT 15  
 AAB34768  
 ID AAB34768 standard; Protein; 272 AA.  
 AC AAB34768;  
 XX  
 DT 26-JAN-2001 (first entry)  
 XX Human secreted protein fragment encoded by DNA clone vq10.1.  
 DE  
 XX Secreted protein; human; autoimmune disorder; multiple sclerosis; ulcer;  
 KW systemic lupus erythematosus; rheumatoid arthritis; anaemia; stroke;  
 KW haematopoiesis regulation; tissue regrowth; wound healing; haemophilia;

KW Alzheimer's disease; Parkinson's disease; Shy-drager syndrome; cancer;  
 KW contraceptive; infection; growth inhibition; hyperproliferative disorder;  
 KW psoriasis.  
 OS Homo sapiens.  
 PN W020005375-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 17-MAR-2000; 2000WO-US07285.  
 XX  
 PR 17-MAR-1999; 99US-0124808.  
 PR 17-MAR-1999; 99US-0124916.  
 PR 17-AUG-1999; 99US-0148639.  
 PR 01-OCT-1999; 99US-0157247.  
 PR 29-NOV-1999; 99US-0167824.  
 PR 15-FEB-2000; 2000US-0182711.  
 XX  
 PA (ALPH-) ALPHAGENE INC.  
 PI Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;  
 DR WPI; 2000-638211/61.  
 DR N-PSDB; AAC59831.  
 XX  
 PT Novel proteins and polypeptides useful for the treatment of e.g.  
 PT multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,  
 PT cancer, Alzheimer's disease, Parkinson's disease, stroke, anemia and  
 PT ulcers.  
 XX Disclosure; Page 485-486; 493pp; English.  
 PS  
 XX This invention relates to 59 human secreted proteins and the nucleotide  
 CC sequences encoding them. Sequences AAC59788-C59846 and AAB34687-B34745  
 CC represent the proteins and their encoding nucleotide sequences, and  
 CC sequences AAB34746-B34771 represent fragments of the proteins. Probes  
 CC for the DNA sequences are represented by sequences AAC59847-C59956. The  
 CC proteins exhibit neuroprotective; dermatological, immunosuppressive,  
 CC antiinflammatory, antiaiemic, nootropic, antiparkinsonian,  
 CC cardioprotective, haemostatic, vulnerary, cytotactic, antipsoptic,  
 CC antibacterial, virucide, and fungicide activity. The proteins and  
 CC nucleotide sequences are useful as nutritional sources or supplements  
 CC and in research. The proteins are useful for treating immune deficiency  
 CC and disorders, which may be genetic or resulting from infections,  
 CC autoimmune disorders such as multiple sclerosis, systemic lupus  
 CC erythematosus, rheumatoid arthritis, and for treating myeloid or lymphoid  
 CC cell deficiencies such as anaemias by regulating haematopoiesis. The  
 CC proteins are also useful in compositions for bone, cartilage, tendon,  
 CC ligament and/or nerve tissue growth or regeneration, for wound healing,  
 CC tissue repair and replacement and in the treatment of wounds, incisions  
 CC and ulcers. Other uses include in the treatment of central and  
 CC peripheral nervous system and neuropathies such as Alzheimer's and  
 CC Parkinson's diseases and Shy-Drager syndrome, and mechanical and  
 CC traumatic disorders, such as spinal cord disorders, head trauma and  
 CC stroke. The proteins may also be used as a contraceptive, and for  
 CC treating coagulation disorders such as haemophilias. The protein and  
 CC nucleotide sequences with cadherin activity are useful for treating  
 CC cancer. Other uses for the protein include for inhibiting the growth,  
 CC infection or function of, or killing, infectious agents such as bacteria,  
 CC virus, fungi and other parasites, for effecting bodily characteristics  
 CC such as height, weight, hair colour, effecting biorhythms or cardiac  
 CC cycles or rhythms, effecting metabolism, catabolism, anabolism,  
 CC processing, utilization, storage, or elimination of dietary fat, lipid,  
 CC protein, carbohydrate, vitamins, minerals, cofactors, effecting  
 CC behavioural characteristics, providing anesthetic effects and for treating  
 CC hyperproliferative disorders such as psoriasis.  
 CC  
 XX Sequence 272 AA;  
 SQ  
 Query Match 98.5%; Score 407; DB 21; Length 272;  
 Best Local Similarity 98.7%; Pred. No. 3.1e-43;  
 Matches 78; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	FPIPLPYCWL	CRALIKRIQAMI	PKGALRVVAQ	QCRVPELV	VAGGICQCL	LAERYSV	ILLDT	60
Db	96	FPIPLPYCWL	CRALIKRIQAMI	PKGALRVVAQ	QCRVPELV	VAGGICQCL	LAERYSV	ILLDT	155
Qy	61	LIGRMLPOL	VCRLVLRCSM	79					
Db	156	LIGRMLPOL	VCRLVLRCSM	174					

Search completed: December 3, 2003, 15:49:53  
Job time : 74.0884 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: December 3, 2003, 15:47:35 ; Search time 25.2585 Seconds  
(without alignments)  
132.334 Million cell updates/sec

Title: US-09-788-308d-2

Perfect score: 413  
Sequence: 1 FPIPLPYCMLCRALIRKRIQA.....TLGRMLPOLVCRLVLRCSM 79

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/aa/6C\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	413	100.0	381	2	US-09-193-877-2 Sequence 2, Appl
2	407	98.5	181	3	US-08-848-580-12 Sequence 12, Appl
3	401	97.1	79	3	US-08-596-684F-6 Sequence 6, Appl
4	401	97.1	257	3	US-08-596-684F-7 Sequence 7, Appl
5	142	34.4	25	4	US-09-019-346A-1 Sequence 1, Appl
6	130	31.5	26	1	US-08-296-898-1 Sequence 1, Appl
7	130	31.5	26	1	US-08-435-019-2 Sequence 2, Appl
8	108	26.2	26	1	US-08-435-019-3 Sequence 3, Appl
9	101	24.5	31	4	US-09-019-346A-2 Sequence 2, Appl
10	91	22.0	80	1	US-08-100-247-3 Sequence 2, Appl
11	91	22.0	80	1	US-08-483-146A-3 Sequence 3, Appl
12	91	22.0	80	1	US-08-232-513A-4 Sequence 4, Appl
13	91	22.0	80	1	US-08-484-594A-3 Sequence 3, Appl
14	91	22.0	80	4	US-09-076-258A-3 Sequence 3, Appl
15	91	22.0	523	1	US-08-100-247-2 Sequence 2, Appl
16	91	22.0	523	1	US-08-483-146A-2 Sequence 2, Appl
17	91	22.0	523	1	US-08-232-513A-3 Sequence 3, Appl
18	91	22.0	523	1	US-08-484-594A-2 Sequence 2, Appl
19	91	22.0	523	4	US-09-352-258A-2 Sequence 2, Appl
20	89	21.5	21	1	US-08-296-898-2 Sequence 2, Appl
21	89	21.5	21	1	US-08-435-019-5 Sequence 5, Appl
22	89	21.5	21	1	US-08-435-019-6 Sequence 6, Appl
23	86	20.8	18	1	US-08-435-019-1 Sequence 1, Appl
24	85	20.6	21	1	US-08-435-019-17 Sequence 17, Appl
25	84	20.3	16	3	US-08-445-422-1 Sequence 1, Appl
26	76	18.4	15	3	US-08-445-422-1 Sequence 1, Appl
27	76	18.4	15	4	US-09-435-204-1 Sequence 1, Appl

28	76	18.4	31	4	US-09-019-346A-3 Sequence 3, Appl
29	68	16.5	13	1	US-08-435-019-10 Sequence 10, Appl
30	66	16.0	16	1	US-08-435-019-8 Sequence 8, Appl
31	64.5	15.6	78	2	US-08-732-228-1 Sequence 1, Appl
32	63.5	15.4	365	2	US-08-222-617A-13 Sequence 13, Appl
33	63.5	15.4	3712	2	US-08-222-617A-4 Sequence 4, Appl
34	63.5	15.4	3712	2	US-08-222-617A-25 Sequence 25, Appl
35	63	15.3	13	1	US-08-435-019-4 Sequence 4, Appl
36	62	15.0	16	1	US-08-435-019-9 Sequence 9, Appl
37	62	15.0	599	4	US-09-252-991A-29067 Sequence 29067, A
38	60.5	14.6	307	4	US-09-252-991A-17664 Sequence 17664, A
39	60.5	14.6	776	4	US-09-252-991A-30380 Sequence 30380, A
40	58.5	14.2	251	4	US-09-252-991A-28922 Sequence 28922, A
41	58	14.0	224	4	US-09-280-598A-49 Sequence 49, Appl
42	58	14.0	277	4	US-09-328-352-4562 Sequence 4562, Ap
43	58	14.0	534	3	US-08-866-181A-5 Sequence 5, Appl
44	58	14.0	535	1	US-07-906-349A-10 Sequence 10, Appl
45	58	14.0	535	1	US-08-167-035-10 Sequence 10, Appl

## ALIGNMENTS

RESULT 1  
US-09-193-877-2  
Sequence 2, Application US/09193877  
Patent No. 5993809  
GENERAL INFORMATION:  
APPLICANT: Weaver, Timothy E.  
TITLE OF INVENTION: LYSOZYME FUSION PROTEINS IN INFECTIONS  
FILE REFERENCE: CMC-127-232  
CURRENT APPLICATION NUMBER: US/09/193,877  
CURRENT FILING DATE: 1998-11-18  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 381  
TYPE: PRT  
ORGANISM: HUMAN  
FEATURE:  
NAME/KEY: PROPER  
LOCATION: (0)...(0)  
US-09-193-877-2

Query Match 100.0%; Score 413; DB 2; Length 381;  
Best Local Similarity 100.0%; Pred. No. 2.3e-45;  
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPIPLPYCMLCRALIRKRIQAIPKGLRVAVAOVCVPLVAGICOCCLAERYSVILDT 60  
DB 201 FPIPLPYCMLCRALIRKRIQAIPKGLRVAVAOVCVPLVAGICOCCLAERYSVILDT 260  
QY 61 TLGRMLPOLVCRLVLRCSM 79  
DB 261 TLGRMLPOLVCRLVLRCSM 279

RESULT 2  
US-08-848-580-12  
Sequence 12, Application US/08848580  
Patent No. 6013619  
GENERAL INFORMATION:  
APPLICANT: Cochran, Charles G  
APPLICANT: Revak, Susan D  
TITLE OF INVENTION: NOVEL PULMONARY SURFACTANTS AND  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: THE SCRIPPS RESEARCH INSTITUTE  
STREET: 10550 No. 6013619th Torrey Pines Road, TPC-8  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA



ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/848,580  
FILING DATE: 28-APR-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/488,123  
FILING DATE: 06-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/419,824  
FILING DATE: 11-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/060,833  
FILING DATE: 12-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/715,397  
FILING DATE: 14-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/293,201  
FILING DATE: 04-JAN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/141,200  
FILING DATE: 06-JAN-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitting, Thomas  
REGISTRATION NUMBER: 34,163  
REFERENCE/DOCKET NUMBER: TSRI 147.5  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-784-2937  
TELEFAX: 619-784-9399  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 181 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-848-580-12

Query Match 98.5%; Score 407; DB 3; Length 181;  
Best Local Similarity 98.7%; Pred. No. 5,8e-45;  
Matches 78; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PPIPLPYCWLGRALIKRIQAMIPKCALRVAVAVQCRVPLVAGGICCCLAERYSVILLDT 60  
Db 1 PPIPLPYCWLGRALIKRIQAMIPKCALRVAVAVQCRVPLVAGGICCCLAERYSVILLDT 60

Qy 61 LGRMLPOLVCRVLVRCSM 79  
Db 61 LGRMLPOLVCRVLVRCSM 79

RESULT 3  
US-08-596-684F-6  
Sequence 6, Application US/08596684F  
Patent No. 6031075  
GENERAL INFORMATION:  
APPLICANT: Weaver, Timothy  
TITLE OF INVENTION: MATURE ALVEOLAR SP-B AND A  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WOOD, HERRON & EVANS, L.L.P.  
STREET: 2700 CAREW TOWER  
CITY: CINCINNATI  
STATE: OH  
COUNTRY: USA  
ZIP: 45202-2917  
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/596,684F  
FILING DATE: 05-FEB-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Josephic, David J  
REGISTRATION NUMBER: 22,849  
REFERENCE/DOCKET NUMBER: CMC-102  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 513-241-2324  
TELEFAX: 513-421-7269  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 79 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 1...79  
OTHER INFORMATION:  
US-08-596-684F-6

Query Match 97.1%; Score 401; DB 3; Length 79;  
Best Local Similarity 98.7%; Pred. No. 1.3e-44;  
Matches 77; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 PPIPLPYCWLGRALIKRIQAMIPKCALRVAVAVQCRVPLVAGGICCCLAERYSVILLDTL 61  
Db 2 PPIPLPYCWLGRALIKRIQAMIPKCALRVAVAVQCRVPLVAGGICCCLAERYSVILLDTL 61

Qy 62 LGRMLPOLVCRVLVRCSM 79  
Db 62 LGRMLPOLVCRVLVRCSM 79

RESULT 4  
US-08-596-684F-7  
Sequence 7, Application US/08596684F  
Patent No. 6031075  
GENERAL INFORMATION:  
APPLICANT: Weaver, Timothy  
TITLE OF INVENTION: MATURE ALVEOLAR SP-B AND A  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WOOD, HERRON & EVANS, L.L.P.  
STREET: 2700 CAREW TOWER  
CITY: CINCINNATI  
STATE: OH  
COUNTRY: USA  
ZIP: 45202-2917  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/596,684F  
FILING DATE: 05-FEB-1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Josephic, David J.  
REGISTRATION NUMBER: 22,849  
REFERENCE/DOCKET NUMBER: CMC-102  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 513-241-2324  
TELEFAX: 513-421-7269  
TELEX:  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 257 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: N-terminal  
FEATURE:  
NAME/KEY: Signal Sequence  
LOCATION: 1...178  
OTHER INFORMATION:  
US-08-596-684F-7

Query Match 97.1%; Score 401; DB 3; Length 257;  
Best Local Similarity 98.7%; Pred. No. 5.2e-44;  
Matches 77; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PIPPLPYCWLGRALIKRIQAMIPKGLRVNAOVCRVPLVAGGICCCLAERYSVIILDTL 61  
DB 180 PIPPLPYCWLGRALIKRIQAMIPKGLRVNAOVCRVPLVAGGICCCLAERYSVIILDTL 239

QY 62 LGRMLPOLVCRVLVLRCSM 79  
DB 240 LGRMLPOLVCRVLVLRCSM 257

RESULT 5  
US-09-019-346A-1  
Sequence 1, Application US/09019346A  
Patent No. 6372720  
GENERAL INFORMATION:  
APPLICANT: Longmuit, Kenneth J.  
APPLICANT: Waring, Alan J.  
APPLICANT: Haynes, Sherry M.  
TITLE OF INVENTION: Liposome Fusion and Delivery Vehicle  
FILE REFERENCE: A65189/RFT/DAV  
CURRENT APPLICATION NUMBER: US/09/019,346A  
CURRENT FILING DATE: 1998-02-05  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 25  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: synthetic  
US-09-019-346A-1

Query Match 34.4%; Score 142; DB 4; Length 25;  
Best Local Similarity 100.0%; Pred. No. 9e-12;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PIPPLPYCWLGRALIKRIQAMIPKG 25  
DB 1 PIPPLPYCWLGRALIKRIQAMIPKG 25

RESULT 6  
US-08-296-898-1  
Sequence 1, Application US/08296898  
Patent No. 5547937  
GENERAL INFORMATION:  
APPLICANT: Diaon, M.

APPLICANT: Lundell, E.  
APPLICANT: Sarin, V.  
APPLICANT: Baxter, C.  
APPLICANT: Absolom, D.  
TITLE OF INVENTION: Pulmonary Surfactant Protein Fragments  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Donald O. Nickey  
ADDRESSEE: Ross Products Division  
ADDRESSEE: Abbott Laboratories  
STREET: 625 Cleveland Avenue  
CITY: Columbus  
STATE: Ohio  
COUNTRY: United States of America  
ZIP: 43215  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS Version 6.21  
SOFTWARE: Wordperfect Version 6.0a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/296,898  
FILING DATE: 29-AUG-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/866,916  
FILING DATE: 10-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Donald O. Nickey  
REGISTRATION NUMBER: 29,092  
REFERENCE/DOCKET NUMBER: 5133 US.D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (614) 624-7080  
TELEFAX: (614) 624-3074  
TELEX: No. 5547937e  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acid residues  
TYPE: Amino acid  
STRANDEDNESS: Single  
TOPOLOGY: linear  
MOLECULE TYPE: Polypeptide  
FEATURE:  
NAME/KEY: SP-B (53-78)  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION: Surfactant polypeptide; fragment 53 to 78  
OTHER INFORMATION: of surfactant-active protein SP-B.  
US-08-296-898-1

Query Match 31.5%; Score 130; DB 1; Length 26;  
Best Local Similarity 100.0%; Pred. No. 3.3e-10;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 YSVIILDTLLGRMLPOLVCRVLVRC 78  
DB 1 YSVIILDTLLGRMLPOLVCRVLVRC 26

RESULT 7  
US-08-435-019-2  
Sequence 2, Application US/08435019  
Patent No. 5753621  
GENERAL INFORMATION:  
APPLICANT: Diaon, Madhup K.  
APPLICANT: Lundell, Edwin O.  
APPLICANT: Sarin, Virender K.  
APPLICANT: Baxter, Constance H.  
TITLE OF INVENTION: Pulmonary Surfactant Protein Fragments  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Donald O. Nickey

STREET: Abbott Laboratories, D-377 AP6D  
CITY: Abbott Park  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60604  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 500 Kb storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS Version 3.30  
SOFTWARE: Wordperfect Version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/435.019  
FILING DATE: 04-MAY-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/866,916  
FILING DATE: 10-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Nickey, Donald O.  
REGISTRATION NUMBER: 29092  
REFERENCE/DOCKET NUMBER: 5133.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (614) 624-7080  
TELEFAX: (614) 624-3074  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acid residues  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: polypeptide  
HYPOTHETICAL: no  
ANTI-SENSE:  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
ORGANISM:  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
HAPLOTYPE:  
TISSUE TYPE:  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
IMMEDIATE SOURCE:  
LIBRARY:  
CLONE:  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT:  
MAP POSITION:  
UNITS:  
FEATURE:  
NAME/KEY: SP-B(53-78)  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION: Surfactant polypeptide;  
OTHER INFORMATION: Fragment 53-78 of surfactant-active protein  
OTHER INFORMATION: known as SP-B, SAP-B, SAP-(Phe), SAP-6 (Phe)  
OTHER INFORMATION: or SPL-(Phe)  
PUBLICATION INFORMATION:  
AUTHORS:  
TITLE:  
JOURNAL:  
VOLUME:  
ISSUE:  
PAGES:  
DATE:  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO:  
US-08-435-019-2

Query Match 31.5%; Score 130; DB 1; Length 26;  
Best Local Similarity 100.0%; Pred. No. 3,3e-10;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 53 YSVILDTLLGRMLPOLVCRVLVLRCS 78  
DB 1 YSVILDTLLGRMLPOLVCRVLVLRCS 26  
RESULT 8  
US-08-435-019-3  
Sequence 3, Application US/08435019  
Patent No. 5753621  
GENERAL INFORMATION:  
APPLICANT: Dhacn, Madhup K.  
APPLICANT: Lundell, Edwin O.  
APPLICANT: Sarin, Virender K.  
APPLICANT: Baxter, Constance H.  
APPLICANT: Absolom, Daryl R.  
TITLE OF INVENTION: Pulmonary Surfactant Protein Fragments  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Donald O. Nickey  
STREET: Abbott Laboratories, D-377 AP6D  
CITY: Abbott Park  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60604  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 500 Kb storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS Version 3.30  
SOFTWARE: Wordperfect Version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/435.019  
FILING DATE: 04-MAY-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/866,916  
FILING DATE: 10-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Nickey, Donald O.  
REGISTRATION NUMBER: 29092  
REFERENCE/DOCKET NUMBER: 5133.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (614) 624-7080  
TELEFAX: (614) 624-3074  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acid residues  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: polypeptide  
HYPOTHETICAL: no  
ANTI-SENSE:  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
ORGANISM:  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
HAPLOTYPE:  
TISSUE TYPE:  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
IMMEDIATE SOURCE:  
LIBRARY:  
CLONE:  
POSITION IN GENOME:

CHROMOSOME/SEGMENT:  
MAP POSITION:  
UNITS:  
FEATURE:  
NAME/KEY: SP-B(53-78)diacm  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION: Surfactant polypeptide;  
OTHER INFORMATION: Fragment 53-78 of surfactant-active protein  
OTHER INFORMATION: known as SP-B, SAP-B, SAP-(Phe), SAP-6 (Phe)  
OTHER INFORMATION: or SPL-(Phe); modified at positions 71 and 77  
OTHER INFORMATION: by providing each of the cysteine residues at  
OTHER INFORMATION: these positions with an S-acetamidomethyl (AcM)  
PUBLICATION INFORMATION:  
AUTHORS:  
TITLE:  
JOURNAL:  
VOLUME:  
ISSUE:  
PAGES:  
DATE:  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO:  
US-08-435-019-3

Query Match 26.2%; Score 108; DB 1; Length 26;  
Best Local Similarity 92.3%; Pred. No. 2.3e-07;  
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 53 YSVILDTLGRMLPOLVCRVLRLCS 78  
DB 1 YSVILDTLGRMLPOLVCRVLRLXS 26

RESULT 9  
US-09-019-346A-2  
Sequence 2, Application US/09019346A  
Patent No. 6372720  
GENERAL INFORMATION:  
APPLICANT: Longmuller, Kenneth J.  
APPLICANT: Waring, Alan J.  
APPLICANT: Haynes, Sherry M.  
TITLE OF INVENTION: Liposome Fusion and Delivery Vehicle  
FILE REFERENCE: A65189/RT/DAY  
CURRENT APPLICATION NUMBER: US/09/019,346A  
CURRENT FILING DATE: 1998-02-05  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 31  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: synthetic  
US-09-019-346A-2

Query Match 24.5%; Score 101; DB 4; Length 31;  
Best Local Similarity 70.8%; Pred. No. 2.3e-06;  
Matches 17; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 PIPYPCWLCRALIKRIQAMIPKG 25  
DB 8 PLPIPCWLAHLIHHIQAIPKG 31

RESULT 10  
US-08-100-247-3  
Sequence 3, Application US/08100247  
Patent No. 5571787  
GENERAL INFORMATION:

APPLICANT: O'BRIEN, JOHN S.  
APPLICANT: KISHIMOTO, YASUO  
TITLE OF INVENTION: PROSAPOSIN AS A NEUROTROPHIC FACTOR  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: KNOBBE, MARTENS, OLSON AND BEAR  
STREET: 620 NEWPORT CENTER DRIVE SIXTEENTH FLOOR  
CITY: NEWPORT BEACH  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/100,247  
FILING DATE: 19930730  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Israel, Ned A.  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: O'BRIEN, 002A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-235-8550  
TELEFAX: 619-235-0176  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
IMMEDIATE SOURCE:  
CLONE: SAPOSIN C  
US-08-100-247-3

Query Match 22.0%; Score 91; DB 1; Length 80;  
Best Local Similarity 27.6%; Pred. No. 0.00013;  
Matches 21; Conservative 15; Mismatches 36; Indels 4; Gaps 1;

OY 7 YCWLCPALIRIAMIPIKG---ALRVAAQVCRVPLVAGICQCLAEYVSVLLDTLL 62  
DB 4 YCEVCEFLVEYVKLIDNNKTEKEILDAPDKMSKLPKSLSECCQEVVDYGSILSL 63

OY 63 GRMLPOLVCRVLRLCS 78  
DB 64 EAVSPFLVCSMLHCS 79

RESULT 11  
US-08-483-146A-3  
Sequence 3, Application US/08483146A  
Patent No. 5696080  
GENERAL INFORMATION:  
APPLICANT: O'Brien, John S.  
APPLICANT: Kishimoto, Yasuo  
TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS  
TITLE OF INVENTION: COMPRISING PROSAPOSIN AND NEUROTROPHIC PEPTIDES DERIVED THEREFROM  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson and Bear  
STREET: 620 Newport Center Blvd. 16th Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483.146A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelien, Ned A  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: MYELOS.002DV1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-235-8550  
TELEFAX: 619-235-0176  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-483-146A-3

Query Match 22.0%; Score 91; DB 1; Length 80;  
Best Local Similarity 27.6%; Pred. No. 0.00013;  
Matches 21; Conservative 15; Mismatches 36; Indels 4; Gaps 1;

Qy 7 YCWLGRALIKRIQAMIPKG---ALRVAVAGVCRVPLVAGGICQCLAEYSVILDTLL 62  
Db 4 YCEVCEFLVKEVTKLIDNNKTEKEILDAPDKMCKSLPKSLSECEQVVDVIGSSITLTL 63  
Qy 63 GRMLPOLVCRVLNRC 78  
Db 64 EEVSPELVCSMLHCS 79

RESULT 12  
US-08-232-513A-4  
Sequence 4, Application US/08232513A  
Patent No. 5700909  
GENERAL INFORMATION:  
APPLICANT: O'Brien, John S.  
TITLE OF INVENTION: Proasoposin and Cytokine-Derived Peptides  
TITLE OF INVENTION: as Therapeutic Agents  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Florea LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232.513A  
FILING DATE: 21-APR-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/100.247  
FILING DATE: 30-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-UD 1643  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1..80  
OTHER INFORMATION: /label= Saposin\_C  
US-08-232-513A-4

Query Match 22.0%; Score 91; DB 1; Length 80;  
Best Local Similarity 27.6%; Pred. No. 0.00013;  
Matches 21; Conservative 15; Mismatches 36; Indels 4; Gaps 1;

Qy 7 YCWLGRALIKRIQAMIPKG---ALRVAVAGVCRVPLVAGGICQCLAEYSVILDTLL 62  
Db 4 YCEVCEFLVKEVTKLIDNNKTEKEILDAPDKMCKSLPKSLSECEQVVDVIGSSITLTL 63  
Qy 63 GRMLPOLVCRVLNRC 78  
Db 64 EEVSPELVCSMLHCS 79

RESULT 13  
US-08-484-594A-3  
Sequence 3, Application US/08484594A  
Patent No. 5714459  
GENERAL INFORMATION:  
APPLICANT: O'Brien, John S.  
APPLICANT: Kishimoto, Yasuo  
TITLE OF INVENTION: USE OF PROSAPOSIN AND NEUTROTROPIC PEPTIDES  
TITLE OF INVENTION: DERIVED THEREFROM  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson and Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484.594A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/100.247  
FILING DATE: 30-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelien, Ned A  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: MYELOS.002DV2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-235-8550  
TELEFAX: 619-235-0176  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 80 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 3, 2003, 15:50:00 ; Search time 47.8299 Seconds

(Without alignments)  
307.186 Million cell updates/sec

Title: US-09-788-308D-2

Perfect score: 1 FPIPLPYCMLCRALIRKIQ.....TLGRMLPOLVCRVLVRCSM 79

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	413	100.0	79	11	US-09-788-308D-2
2	407	98.5	181	8	US-08-488-123-12
3	407	98.5	381	12	US-10-236-031B-10
4	140	33.9	154	9	US-09-925-302-458
5	91	22.0	80	9	US-09-767-007A-3
6	91	22.0	80	10	US-09-753-126-3
7	91	22.0	479	11	US-09-978-418-40
8	91	22.0	523	9	US-09-767-007A-2
9	91	22.0	524	10	US-09-870-753-60
10	91	22.0	524	12	US-09-751-708A-60
11	91	22.0	527	9	US-09-870-759-61
12	91	22.0	527	12	US-09-751-708A-61
13	91	22.0	537	15	US-10-060-036-73
14	91	22.0	592	10	US-09-753-126-4
15	87	21.1	554	12	US-10-203-194-176

16	77.5	18.8	399	12	US-10-094-749-2219	Sequence 2219, Ap
17	73.5	17.8	521	12	US-10-276-162-1	Sequence 1, Appl
18	71.5	17.3	400	12	US-10-017-161-2290	Sequence 2290, Ap
19	66	16.0	318	14	US-10-042-417-42	Sequence 42, Appl
20	63	15.3	283	15	US-10-156-761-12760	Sequence 12760, A
21	62.5	15.1	457	12	US-09-826-509-579	Sequence 579, App
22	62.5	15.1	457	15	US-10-225-567A-469	Sequence 469, App
23	60.5	14.6	324	15	US-10-156-761-14083	Sequence 14083, A
24	59.5	14.4	202	11	US-09-989-442-103	Sequence 103, App
25	59.5	14.4	260	9	US-09-761-288-35	Sequence 35, Appl
26	59.5	14.4	260	9	US-09-761-288-36	Sequence 36, Appl
27	59.5	14.4	260	11	US-09-898-586-35	Sequence 35, Appl
28	59.5	14.4	260	11	US-09-898-586-36	Sequence 36, Appl
29	59.5	14.4	272	12	US-09-907-218-35	Sequence 35, Appl
30	59.5	14.4	304	10	US-09-747-835A-58	Sequence 58, Appl
31	59.5	14.4	310	9	US-09-761-288-4	Sequence 4, Appl
32	59.5	14.4	310	9	US-09-761-288-38	Sequence 38, Appl
33	59.5	14.4	310	9	US-09-761-288-67	Sequence 67, Appl
34	59.5	14.4	310	9	US-09-761-288-89	Sequence 89, Appl
35	59.5	14.4	310	11	US-09-898-586-38	Sequence 38, Appl
36	59.5	14.4	310	11	US-09-898-586-67	Sequence 67, Appl
37	59.5	14.4	310	11	US-09-898-586-89	Sequence 89, Appl
38	59.5	14.4	310	12	US-09-907-218-33	Sequence 33, Appl
39	59.5	14.4	310	12	US-09-965-422-65	Sequence 65, Appl
40	59.5	14.4	313	11	US-09-791-932-63	Sequence 63, Appl
41	59.5	14.4	417	9	US-09-741-669-358	Sequence 358, App
42	59.5	14.4	466	12	US-10-419-190-2	Sequence 2, Appl
43	59.5	14.3	165	12	US-10-094-749-3036	Sequence 3036, Ap
44	59	14.3	15	12	US-10-094-749-3036	Sequence 3036, Ap
45	58.5	14.2	295	9	US-09-815-242-5134	Sequence 5134, Ap

## ALIGNMENTS

RESULT 1  
US-09-788-308D-2  
; Sequence 2, Application US/09788308D  
; Publication No. US20030040468A1  
; GENERAL INFORMATION:  
; APPLICANT: No. US20030040468A1, Western University  
; TITLE OF INVENTION: Polypeptide Pulmonary Surfactants  
; FILE REFERENCE: 6374  
; CURRENT APPLICATION NUMBER: US/09/788, 308D  
; PRIOR FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: US 60/182, 847  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 79  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-788-308D-2

Query Match 100.0%; Score 413; DB 11; Length 79;  
Best Local Similarity 100.0%; Pred. No. 7.5e-43;  
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FPIPLPYCMLCRALIRKIQAMIPKGLRVAVQVCRVPLVAGICCOCLAERYVILDT 60  
DB 1 FPIPLPYCMLCRALIRKIQAMIPKGLRVAVQVCRVPLVAGICCOCLAERYVILDT 60

QY 61 LLGRMLPOLVCRVLVRCSM 79  
DB 61 LLGRMLPOLVCRVLVRCSM 79

RESULT 2  
US-08-488-123-12  
; Sequence 12, Application US/08488123  
; Publication No. US20030099696A1  
; GENERAL INFORMATION:

APPLICANT: Cochran, Charles G  
APPLICANT: Revak, Susan D  
TITLE OF INVENTION: PEPTIDE-CONTAINING LIPOSOMAL SURFACTANT  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESSES:  
ADDRESS: The Scripps Research Clinic, Office of Patent  
ATTORNEY: Counsel  
STREET: 10666 No. US20030099696A1ch Torrey Pines Road, TPC8  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,123  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/419,824  
FILING DATE: 11-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/060,833  
FILING DATE: 12-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/715,397  
FILING DATE: 14-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/293,201  
FILING DATE: 04-JAN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 141,200  
FILING DATE: 06-JAN-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Logan, April C  
REGISTRATION NUMBER: 33,950  
REFERENCE/DOCKET NUMBER: SCR1738P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-554-2937  
TELEFAX: 619-554-6312  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 181 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-488-123-12

Query Match 98.5%; Score 407; DB 8; Length 181;  
Best Local Similarity 98.7%; Pred. No. 1e-41;  
Matches 78; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPIPLPYCMLCRALIKRIQAMIPKCALRVAAVQVCRVPLVAGGICQCLAERYSVILDT 60  
DB 1 FPIPLPYCMLCRALIKRIQAMIPKCALRVAAVQVCRVPLVAGGICQCLAERYSVILDT 60

QY 61 LLGRMLPOLVCRVLVRCGM 79  
DB 61 LLGRMLPOLVCRVLVRCGM 79

RESULT 3  
US-10-236-031B-10  
Sequence 10, Application US/10236031B  
Publication No. US20030219760A1  
GENERAL INFORMATION:  
APPLICANT: Gordon, Gavin J.  
APPLICANT: Jensen, Roderick V.  
APPLICANT: Guillan, Steven R.  
APPLICANT: Bueno, Raphael

TITLE OF INVENTION: Diagnostic and Prognostic Tests  
FILE REFERENCE: B00801/70265 (JRV/JAV)  
CURRENT APPLICATION NUMBER: US/10/236,031B  
CURRENT FILING DATE: 2002-09-05  
PRIOR APPLICATION NUMBER: US 60/317,389  
PRIOR FILING DATE: 2001-09-05  
PRIOR APPLICATION NUMBER: US 60/407,431  
PRIOR FILING DATE: 2002-08-30  
NUMBER OF SEQ ID NOS: 102  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 10  
LENGTH: 381  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-236-031B-10

Query Match 98.5%; Score 407; DB 12; Length 381;  
Best Local Similarity 98.7%; Pred. No. 2.3e-41;  
Matches 78; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPIPLPYCMLCRALIKRIQAMIPKCALRVAAVQVCRVPLVAGGICQCLAERYSVILDT 60  
DB 201 FPIPLPYCMLCRALIKRIQAMIPKCALRVAAVQVCRVPLVAGGICQCLAERYSVILDT 260

QY 61 LLGRMLPOLVCRVLVRCGM 79  
DB 261 LLGRMLPOLVCRVLVRCGM 279

RESULT 4  
US-09-925-302-458  
Sequence 458, Application US/09925302  
Patent No. US20020044941A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA104  
CURRENT APPLICATION NUMBER: US/09/925,302  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05918  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 896  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 458  
LENGTH: 154  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (111)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (122)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (131)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (132)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-302-458

Query Match 33.9%; Score 140; DB 9; Length 154;  
Best Local Similarity 100.0%; Pred. No. 2.8e-09;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 RYSVILDTLLGRMLPOLVCRVLVRCGM 79  
DB 1 RYSVILDTLLGRMLPOLVCRVLVRCGM 28



```
RESULT 5
US-09-767-007A-3
; Sequence 3, Application US/09767007A
; Patent No. US20020072725A1
; GENERAL INFORMATION:
; APPLICANT: John S. O'Brien
; APPLICANT: Yasuo Kishimoto
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS COMPRISING
; FILE REFERENCE: WELLOS.2DC1C1
; CURRENT FILING DATE: 2001-01-22
; PRIOR FILING DATE: 1997-10-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Artificial peptide
US-09-767-007A-3

Query Match          22.0%; Score 91; DB 9; Length 80;
Best Local Similarity 27.6%; Pred. No. 0.0012;
Matches 21; Conservative 15; Mismatches 36; Indels 4; Gaps 1;

OY 7 YCMLCRALIRIQAMIPKG---ALRYVAQVCRVPLVAGGICQCLAERYVLLDTLL 62
DB 4 YCEVCEFLVKEVTLIDNNKTEKELIDAFDKMCKLPKSLSECOEVDVTGSSILSILL 63

OY 63 GRMLPOLVCRVLRLCS 78
DB 64 EEVSPELVCSMLHLCS 79

RESULT 6
US-09-753-126-3
; Sequence 3, Application US/09753126
; Patent No. US20020127219A1
; GENERAL INFORMATION:
; APPLICANT: OKKELS, JENS SIGURD
; APPLICANT: JENSEN, ANNE DAM
; APPLICANT: HALKIER, TOREBN
; APPLICANT: JENSEN, RIKKE BOLDING
; TITLE OF INVENTION: IMPROVED LYSOSOMAL ENZYMES AND LYSOSOMAL ENZYME
; FILE REFERENCE: 31-000600US
; CURRENT APPLICATION NUMBER: US/09/753,126
; CURRENT FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: PA 1999 01891
; PRIOR FILING DATE: 1999-12-30
; PRIOR APPLICATION NUMBER: 60/174,652
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: PA 200 00865
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/210,984
; PRIOR FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: 60/211,124
; PRIOR FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: PA 2000 01027
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/217,497
; PRIOR FILING DATE: 2000-07-11
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-753-126-3
```

```
Query Match          22.0%; Score 91; DB 10; Length 80;
Best Local Similarity 27.6%; Pred. No. 0.0012;
Matches 21; Conservative 15; Mismatches 36; Indels 4; Gaps 1;

OY 7 YCMLCRALIRIQAMIPKG---ALRYVAQVCRVPLVAGGICQCLAERYVLLDTLL 62
DB 4 YCEVCEFLVKEVTLIDNNKTEKELIDAFDKMCKLPKSLSECOEVDVTGSSILSILL 63

OY 63 GRMLPOLVCRVLRLCS 78
DB 64 EEVSPELVCSMLHLCS 79

RESULT 7
US-09-978-418-40
; Sequence 40, Application US/09978418
; Publication No. US20030118997A1
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephan
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 142.US.REG
; CURRENT APPLICATION NUMBER: US/09/978,418
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: 60/311,305
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/314,734
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/318,204
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/326,470
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: JPatent
; SEQ ID NO 40
; LENGTH: 479
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-978-418-40

Query Match          22.0%; Score 91; DB 11; Length 479;
Best Local Similarity 27.6%; Pred. No. 0.0091;
Matches 21; Conservative 15; Mismatches 36; Indels 4; Gaps 1;

OY 7 YCMLCRALIRIQAMIPKG---ALRYVAQVCRVPLVAGGICQCLAERYVLLDTLL 62
DB 317 YCEVCEFLVKEVTLIDNNKTEKELIDAFDKMCKLPKSLSECOEVDVTGSSILSILL 376

OY 63 GRMLPOLVCRVLRLCS 78
DB 377 EEVSPELVCSMLHLCS 392

RESULT 8
US-09-767-007A-2
; Sequence 2, Application US/09767007A
; Patent No. US20020072725A1
; GENERAL INFORMATION:
; APPLICANT: John S. O'Brien
; APPLICANT: Yasuo Kishimoto
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS COMPRISING
; FILE REFERENCE: WELLOS.2DC1C1
; CURRENT FILING DATE: 2001-01-22
; PRIOR FILING DATE: 1997-10-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 523
; TYPE: PRT
; ORGANISM: homo sapiens
```

US-09-767-007A-2

Query Match 22.0%; Score 91; DB 9; Length 523;  
Best Local Similarity 27.6%; Pred. No. 0.01;  
Matches 21; Conservative 15; Mismatches 36; Indels 4; Gaps 1;

QY 7 YCMLCRALIKRIQMIPIKG---ALRVAAQVCRVPLVAGGICQCLAEYSVILLDTLL 62  
DB 313 YCEVCEFLVKEVTXKIDNNKTEKEILDADFDMKCSKLPKSLSECCQEVVDVTGSSILSTILL 372

QY 63 GRMLPOLYCRVLVLRCS 78  
DB 373 EEVSPBLVCSMLHLCS 388

RESULT 9  
US-09-870-759-60  
; Sequence 60, Application US/09870759  
; Patent No. US20020177551A1  
; GENERAL INFORMATION:

; APPLICANT: TERMAN, David S  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE  
; FILE REFERENCE: 870759  
; CURRENT APPLICATION NUMBER: US/09/870,759  
; CURRENT FILING DATE: 2002-01-14  
; PRIOR APPLICATION NUMBER: US 60/208,128  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 166  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 60  
; LENGTH: 524  
; TYPE: PRT

US-09-870-759-60  
; ORGANISM: Homo sapiens

Query Match 22.0%; Score 91; DB 10; Length 524;  
Best Local Similarity 27.6%; Pred. No. 0.01;  
Matches 21; Conservative 15; Mismatches 36; Indels 4; Gaps 1;

QY 7 YCMLCRALIKRIQMIPIKG---ALRVAAQVCRVPLVAGGICQCLAEYSVILLDTLL 62  
DB 314 YCEVCEFLVKEVTXKIDNNKTEKEILDADFDMKCSKLPKSLSECCQEVVDVTGSSILSTILL 373

QY 63 GRMLPOLYCRVLVLRCS 78  
DB 374 EEVSPBLVCSMLHLCS 389

RESULT 10  
US-09-751-708A-60  
; Sequence 60, Application US/09751708A  
; Publication No. US20030157113A1  
; GENERAL INFORMATION:

; APPLICANT: TERMAN, David S  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE  
; FILE REFERENCE: 751708  
; CURRENT APPLICATION NUMBER: US/09/751,708A  
; CURRENT FILING DATE: 2002-10-15  
; PRIOR APPLICATION NUMBER: US 60/173,371  
; PRIOR FILING DATE: 1999-12-28  
; NUMBER OF SEQ ID NOS: 166  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 60  
; LENGTH: 524  
; TYPE: PRT

US-09-751-708A-60  
; ORGANISM: Homo sapiens

Query Match 22.0%; Score 91; DB 12; Length 524;  
Best Local Similarity 27.6%; Pred. No. 0.01;  
Matches 21; Conservative 15; Mismatches 36; Indels 4; Gaps 1;

QY 7 YCMLCRALIKRIQMIPIKG---ALRVAAQVCRVPLVAGGICQCLAEYSVILLDTLL 62

DB 314 YCEVCEFLVKEVTXKIDNNKTEKEILDADFDMKCSKLPKSLSECCQEVVDVTGSSILSTILL 373

QY 63 GRMLPOLYCRVLVLRCS 78  
DB 374 EEVSPBLVCSMLHLCS 389

RESULT 11  
US-09-870-759-61

; Sequence 61, Application US/09870759,  
; Patent No. US20020177551A1  
; GENERAL INFORMATION:

; APPLICANT: TERMAN, David S  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE  
; FILE REFERENCE: 870759  
; CURRENT APPLICATION NUMBER: US/09/870,759  
; CURRENT FILING DATE: 2002-01-14  
; PRIOR APPLICATION NUMBER: US 60/208,128  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 166  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 61  
; LENGTH: 527  
; TYPE: PRT

US-09-870-759-61  
; ORGANISM: Homo sapiens

Query Match 22.0%; Score 91; DB 10; Length 527;  
Best Local Similarity 27.6%; Pred. No. 0.01;  
Matches 21; Conservative 15; Mismatches 36; Indels 4; Gaps 1;

QY 7 YCMLCRALIKRIQMIPIKG---ALRVAAQVCRVPLVAGGICQCLAEYSVILLDTLL 62  
DB 317 YCEVCEFLVKEVTXKIDNNKTEKEILDADFDMKCSKLPKSLSECCQEVVDVTGSSILSTILL 376

QY 63 GRMLPOLYCRVLVLRCS 78  
DB 377 EEVSPBLVCSMLHLCS 392

RESULT 12  
US-09-751-708A-61

; Sequence 61, Application US/09751708A  
; Publication No. US20030157113A1  
; GENERAL INFORMATION:  
; APPLICANT: TERMAN, David S  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE  
; FILE REFERENCE: 751708  
; CURRENT APPLICATION NUMBER: US/09/751,708A  
; CURRENT FILING DATE: 2002-10-15  
; PRIOR APPLICATION NUMBER: US 60/173,371  
; PRIOR FILING DATE: 1999-12-28  
; NUMBER OF SEQ ID NOS: 166  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 61  
; LENGTH: 527  
; TYPE: PRT

US-09-751-708A-61  
; ORGANISM: Homo sapiens

Query Match 22.0%; Score 91; DB 12; Length 527;  
Best Local Similarity 27.6%; Pred. No. 0.01;  
Matches 21; Conservative 15; Mismatches 36; Indels 4; Gaps 1;

QY 7 YCMLCRALIKRIQMIPIKG---ALRVAAQVCRVPLVAGGICQCLAEYSVILLDTLL 62  
DB 317 YCEVCEFLVKEVTXKIDNNKTEKEILDADFDMKCSKLPKSLSECCQEVVDVTGSSILSTILL 376

QY 63 GRMLPOLYCRVLVLRCS 78  
DB 377 EEVSPBLVCSMLHLCS 392

RESULT 13  
US-10-060-036-73  
; Sequence 73, Application US/10060036  
; Publication No. US20030073144A1  
; GENERAL INFORMATION:  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Persing, David H.  
; APPLICANT: Hepler, William T.  
; APPLICANT: Jiang, Yugu  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.566  
; CURRENT APPLICATION NUMBER: US/10/060,036  
; CURRENT FILING DATE: 2002-01-30  
; NUMBER OF SEQ ID NOS: 4560  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 73  
; LENGTH: 527  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-060-036-73

Query Match 22.0%; Score 91; DB 15; Length 527;  
Best Local Similarity 27.6%; Pred. No. 0.01;  
Matches 21; Conservative 15; Mismatches 36; Indels 4; Gaps 1;

QY 7 YCWLCRALIKRIQAMIPKG---ALRYVAQVCRVPLVVGICQCLAEERYSVLLDPTLL 62  
DB 317 YCEVCEFLVKEVTKLIDNNKTEKEILDAPDKMCKSLPKSISEQEVVDYTGSSILSL 376  
QY 63 GRMLPOLVCRVLVRCSS 78  
DB 377 EEVSPELVCSMLHICS 392

RESULT 14  
US-09-753-126-4  
; Sequence 4, Application US/09753126  
; Patent No. US20020127219A1  
; GENERAL INFORMATION:  
; APPLICANT: OKKELS, JENS SIGURD  
; APPLICANT: JENSEN, ANNE DAM  
; APPLICANT: HALKIER, TORBEN  
; APPLICANT: JENSEN, RIKKE BOLDING  
; TITLE OF INVENTION: IMPROVED LYSOSOMAL ENZYMES AND LYSOSOMAL ENZYME  
; FILE REFERENCE: 31-000600US  
; CURRENT APPLICATION NUMBER: US/09/753,126  
; CURRENT FILING DATE: 2001-06-11  
; PRIOR APPLICATION NUMBER: PA 1999 01891  
; PRIOR FILING DATE: 1999-12-30  
; PRIOR APPLICATION NUMBER: 60/174,652  
; PRIOR FILING DATE: 2000-01-06  
; PRIOR APPLICATION NUMBER: PA 200 00865  
; PRIOR FILING DATE: 2000-06-02  
; PRIOR APPLICATION NUMBER: 60/210,984  
; PRIOR FILING DATE: 2000-06-12  
; PRIOR APPLICATION NUMBER: 60/211,124  
; PRIOR FILING DATE: 2000-06-12  
; PRIOR APPLICATION NUMBER: PA 2000 01027  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: 60/217,497  
; PRIOR FILING DATE: 2000-07-11  
; NUMBER OF SEQ ID NOS: 147  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 592  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Chimeric  
; OTHER INFORMATION: SapC-linker-GCB polypeptide  
US-09-753-126-4

Query Match 22.0%; Score 91; DB 10; Length 592;  
Best Local Similarity 27.6%; Pred. No. 0.011;  
Matches 21; Conservative 15; Mismatches 36; Indels 4; Gaps 1;

QY 7 YCWLCRALIKRIQAMIPKG---ALRYVAQVCRVPLVVGICQCLAEERYSVLLDPTLL 62  
DB 4 YCEVCEFLVKEVTKLIDNNKTEKEILDAPDKMCKSLPKSISEQEVVDYTGSSILSL 63  
QY 63 GRMLPOLVCRVLVRCSS 78  
DB 64 EEVSPELVCSMLHICS 79

RESULT 15  
US-10-205-194-176  
; Sequence 176, Application US/10205194  
; Publication No. US20030134301A1  
; GENERAL INFORMATION:  
; APPLICANT: Warner-Lambert Company  
; APPLICANT: Lee, Kevin  
; APPLICANT: Dixon, Alistair  
; APPLICANT: Brookbank, Robert  
; APPLICANT: Pincock, Robert  
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain  
; FILE REFERENCE: WL-A-018201  
; CURRENT APPLICATION NUMBER: US/10/205,194  
; CURRENT FILING DATE: 5200-07-24  
; PRIOR APPLICATION NUMBER: GB 0118354.0  
; PRIOR FILING DATE: 2001-07-27  
; NUMBER OF SEQ ID NOS: 177  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 176  
; LENGTH: 554  
; TYPE: PRT  
; ORGANISM: Rattus sp.  
; FEATURE:  
; OTHER INFORMATION: Prosaposin  
US-10-205-194-176

Query Match 21.1%; Score 87; DB 12; Length 554;  
Best Local Similarity 23.7%; Pred. No. 0.033;  
Matches 18; Conservative 21; Mismatches 33; Indels 4; Gaps 1;

QY 7 YCWLCRALIKRIQAMIPKG---ALRYVAQVCRVPLVVGICQCLAEERYSVLLDPTLL 62  
DB 313 FCQVQCLVWRKSELINNATELLIKGSKACSLLPAPASTKCOEVLTFFGSLDVL 372  
QY 63 GRMLPOLVCRVLVRCSS 78  
DB 373 HEVNPFLCGVLSLCS 388

Search completed: December 3, 2003, 15:56:08  
Job time : 48.8299 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 3, 2003, 15:45:30 / Search time 23.1088 Seconds  
(without alignments)  
328.763 Million cell updates/sec

Title: US-09-788-308D-2

Sequence: 1 FPIPLPYCWLGRALIKRIDA.....TLIGRLPLQVCRVLKRCM 79

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 9616682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	413	100.0	381 1 LNHUB	pulmonary surfactant
2	350	84.7	2 A29072	pulmonary surfactant
3	344	83.3	370 1 LNRBB	pulmonary surfactant
4	343	83.1	376 2 S02766	pulmonary surfactant
5	332	80.4	359 2 I46531	surfactant protein
6	321	77.7	79 1 LNPGL	pulmonary surfactant
7	210	50.8	60 2 A29667	pulmonary surfactant
8	93.5	22.6	965 2 T00207	P109 protein - sll
9	93	22.5	80 2 S21770	saposin-C - bovine
10	91	22.1	527 1 SAHUP	saposin precursor
11	87	21.1	554 1 A28716	saposin precursor
12	83	20.1	81 2 A32026	glucosylceramide b
13	63.5	15.4	1407 2 S59823	probable membrane
14	63.5	15.4	3712 1 YGCEVC	alpha-aminodipyl-
15	63	15.3	277 2 S44252	raffinose operon r
16	63	15.3	557 1 JH0604	saposin precursor
17	62.5	15.1	92 2 F64452	hypothetical prote
18	62.5	15.1	460 2 JC2194	vasoactive intesti
19	62	15.0	336 2 E96814	hypothetical prote
20	62	15.0	1217 2 T00607	hypothetical prote
21	61.5	14.9	231 2 D82555	heme ABC transport
22	61.5	14.9	277 1 QOBR02	HKRF protein - hu
23	61	14.8	295 2 H64364	NADH2 dehydrogenas
24	61	14.8	374 2 G81926	probable polyamine
25	61	14.8	419 2 C81179	spermidine/putresc
26	61	14.8	566 2 AG0158	protein-Np1-phosph
27	60.5	14.6	223 2 F83262	probable two-compo
28	60.5	14.6	292 2 A75409	carboxymethylenebu
29	60.5	14.6	608 2 S09790	hypothetical prote

RESULT 1

LNHUB

Pulmonary surfactant protein B precursor [validated] - human

N:Alternate names: pulmonary surfactant proteolipid SP-B; pulmonary surfactant-associated

C:Species: Homo sapiens (man)

C>Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 08-Dec-2000

C:Accession: A31361; A28461; A27794; A27592; J00162; S21134

R:Pilot-Matias, T.J.; Kister, S.E.; Fox, J.L.; Kropp, K.; Glasser, S.W.; Whitsett, J.A.

DNA 8, 75-86, 1989

A:Title: Structure and organization of the gene encoding human pulmonary surfactant prot

A:Reference number: A31361; MUID:89170128; PMID:2924687

A:Accession: A31361

A:Molecule type: DNA

A:Residues: 1-381 <PIL>

A:Cross-references: GB:M24461

A>Note: the codon given for residue 131 (ATT) is inconsistent with the authors' translat

A>Note: this protein is encoded by a single gene

R:Jacobs, K.A.; Phelps, D.S.; Steinbrink, R.; Fisch, J.; Kriz, R.; Mitsock, L.; Doughert

J. Biol. Chem. 262, 9808-9811, 1987

A:Title: Isolation of a cDNA clone encoding a high molecular weight precursor to a 6-kDa

A:Reference number: A28461; MUID:87250653; PMID:3597440

A:Accession: A28461

A:Molecule type: mRNA

A:Residues: 1-227, 'A', 229-381 <JAC>

A:Cross-references: GB:U02761; NID:g190673; PIDN:AAA60212.1; PID:g190674

A>Note: part of this sequence, including the amino end of the mature protein, was confir

R:Glasser, S.W.; Korthagen, T.R.; Weaver, T.; Pilot-Matias, T.; Fox, J.L.; Whitsett, J.

Proc. Natl. Acad. Sci. U.S.A. 84, 4007-4011, 1987

A:Title: cDNA and deduced amino acid sequence of human pulmonary surfactant-associated p

A:Reference number: A27794; MUID:87231940; PMID:3035561

A:Accession: A27794

A:Molecule type: mRNA

A:Residues: 'EPR', 99-317, 'L', 319-381 <GLA>

A:Cross-references: GB:M16764; NID:g338410; PIDN:AAA8099.1; PID:g338411

A>Note: 131-ile was also found

A>Note: part of this sequence, including the amino end of the mature protein, was confir

R:Revay, S.D.; Merritt, T.A.; Deglyee, E.; Stefani, L.; Courtney, M.; Hallman, M.; Coch

J. Clin. Invest. 81, 826-833, 1988

A:Title: Use of human surfactant low molecular weight apoproteins in the reconstruction

A:Reference number: A27592; MUID:88139786; PMID:3343343

A:Accession: A27592

A:Molecule type: mRNA

A:Residues: 139-177, 'V', 179-227, 'A', 228-381 <REV>

A:Cross-references: GB:M19097

A>Note: part of this sequence, including the amino end of the mature protein, was confir

A>Note: the mature protein (SP 18) consists of two identical disulfide-bonded 9k polypep

R:Mizumoto, M.; Adachi, H.

Sapporo Igaku Zasshi 56, 731-742, 1987

A:Title: Primary structure of a hydrophobic 6kDa apoprotein (SP6) of human pulmonary su

A:Reference number: J00162

A:Accession: J00162

A:Molecule type: protein

dihydroaeruginosic  
probable PTS syate  
glutamate-5-semial  
gamma-glutamylphos  
NADH2 dehydrogenas  
NADH2 dehydrogenas  
hypothetical prote  
probable C4-dicarb  
NADH2 dehydrogenas  
probable DNA repai  
probable primosoma  
NK-lysin protein -  
thamosyltransfera  
thamosyltransfera  
glutaminase A [imp

A:Residues: 201-207, 'X', 209-210, 'X', 212-227, 'A', 229-234, 'X', 236-245, 'X', 247, 'L', 249-253, R.Johansson, J.; Joernvall, H.; Curedre, T.  
FEBS Lett. 301, 165-167, 1992  
A:Title: Human surfactant polypeptide SP-B. Disulfide bridges, C-terminal end, and pepti  
A:Reference number: S21134; MUID:92233937; PMID:1568474  
A:Accession: S21134  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 201-227, 'I', 229-279 <JOH>  
A:Note: 228-Ala was also found  
C:Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers t  
C:Genetics:  
A:Gene: GDB:SFTPB; SFTPB3; SP-B  
A:Cross-references: GDB:120374; OMIM:178640  
A:Map position: 2p12-2p11.2  
A:Intons: 23/1, 65/3, 89/3, 131/3, 194/3, 224/3, 286/1, 334/3, 361/3  
C:Superfamily: pulmonary surfactant protein B; saposin repeat homology  
C:Keywords: alveolar proteinosis; gaseous exchange; glycoprotein; lipoprotein; lung; pu  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-200/Domain: propeptide #status predicted <PRO>  
F:61-153/Domain: saposin repeat homology <SAP1>  
F:200-287/Domain: saposin repeat homology <SAP2>  
F:201-279/Product: pulmonary surfactant protein B, 9K form #status predicted <SP9>  
F:201-256/Product: pulmonary surfactant protein B, 6K form #status experimental <SP6>  
F:291-376/Domain: saposin repeat homology <SAP3>  
F:69-143, 72-137, 100-112, 299-366, 302-360, 335-335/Disulfide bonds: #status predicted  
F:129, 311/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:208-277, 211-271, 235-246/Disulfide bonds: #status experimental  
F:248/Disulfide bonds: interchain #status experimental

Query Match 100.0%; Score 413; DB 1; Length 381;  
Best Local Similarity 100.0%; Pred. No. 3, 1e-39;  
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAQVCRVPLVAGGICQCLAERYVILLDT 60  
Db 201 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAQVCRVPLVAGGICQCLAERYVILLDT 260

Qy 61 LIGRMLPOLVCRVLVRCSM 79  
Db 261 LIGRMLPOLVCRVLVRCSM 279

RESULT 2  
A29072  
pulmonary surfactant protein SP 18 precursor - dog (fragment)  
C:Species: Canis lupus familiaris (dog)  
C>Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 20-Aug-1999  
C:Accession: B29072; A29072  
R:Hawgood, S.; Benson, B.J.; Schilling, J.; Damm, D.; Clements, J.A.; White, R.T.  
Proc. Natl. Acad. Sci. U.S.A. 84, 66-70, 1987  
A:Title: Nucleotide and amino acid sequences of pulmonary surfactant protein SP 18 and e  
A:Reference number: A29072; MUID:870992398; PMID:3467361  
A:Accession: B29072  
A:Molecule type: mRNA  
A:Residues: 1-363 <HAW>  
A:Cross-references: GB:M5170; NID:G164077; PIDN:AAA30893.1; PID:G164078  
A:Accession: A29072  
A:Molecule type: protein  
A:Residues: 182-210 <HAW>  
C:Superfamily: pulmonary surfactant protein B; saposin repeat homology  
F:1-14/Domain: signal sequence #status predicted <SIG>  
F:15-180/Domain: propeptide #status predicted <PRO>  
F:54-146/Domain: saposin repeat homology <SAP1>  
F:180-267/Domain: saposin repeat homology <SAP2>  
F:181-363/Product: pulmonary surfactant protein SP 18 #status experimental <MAT>  
F:273-358/Domain: saposin repeat homology <SAP3>

Query Match 84.7%; Score 350; DB 2; Length 363;  
Best Local Similarity 84.4%; Pred. No. 4, 1e-32;  
Matches 65; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Qy 2 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAQVCRVPLVAGGICQCLAERYVILLDTL 61

Db 182 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAQVCRVPLVAGGICQCLAERYVILLDTL 241

Qy 62 LIGRMLPOLVCRVLVRC 78  
Db 242 LIGRMLPOLVCRVLVRC 258

RESULT 3  
LNRBB  
pulmonary surfactant protein B precursor - rabbit  
N:Alternate names: pulmonary surfactant-associated protein-B  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C>Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 22-Jun-1999  
C:Accession: A32421  
R:Xu, J.; Richardson, C.; Ford, C.; Spencer, T.; Li-Juan, Y.; Mackie, G.; Hammond, G.;  
Biochem. Biophys. Res. Commun. 160, 335-332, 1989  
A:Title: Isolation and characterization of the cDNA for pulmonary surfactant-associated  
A:Reference number: A32421; MUID:89228033; PMID:2469419  
A:Accession: A32421  
A:Molecule type: mRNA  
A:Residues: 1-370 <XU>  
A:Cross-references: GB:M24901; NID:G165707; PIDN:AAA31466.1; PID:G165708  
A:Note: the authors translated the codon CCG for residue 184 as Arg and CAG for residue  
C:Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers  
C:Superfamily: pulmonary surfactant protein B; saposin repeat homology  
C:Keywords: alveolar proteinosis; gaseous exchange; glycoprotein; lipoprotein; lung; pu  
F:121-184/Domain: signal sequence #status predicted <PRO>  
F:122-184/Domain: propeptide #status predicted <PRO>  
F:62-154/Domain: saposin repeat homology <SAP1>  
F:184-271/Domain: saposin repeat homology <SAP2>  
F:185-263/Product: pulmonary surfactant protein B, 9K form #status predicted <SP9>  
F:185-263/Product: pulmonary surfactant protein B, 6K form #status predicted <SP6>  
F:280-365/Domain: saposin repeat homology <SAP3>  
F:300/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 83.3%; Score 344; DB 1; Length 370;  
Best Local Similarity 82.1%; Pred. No. 2e-31;  
Matches 64; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy 1 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAQVCRVPLVAGGICQCLAERYVILLDT 60  
Db 185 FPIPLPYCWLGRALIKRIQAMIPKGLRVAVAQVCRVPLVAGGICQCLAERYVILLDT 244

Qy 61 LIGRMLPOLVCRVLVRC 78  
Db 245 LIGRMLPOLVCRVLVRC 262

RESULT 4  
S02766  
pulmonary surfactant protein B precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 20-Aug-1999  
C:Accession: S02766  
R:Emile, P.A.; Shannon, J.M.; Mason, R.J.; Fisher, J.H.  
Biochem. Biophys. Acta 994, 215-221, 1989  
A:Title: cDNA and deduced amino acid sequence for the rat hydrophobic pulmonary surfact  
A:Reference number: S02766; MUID:89150284; PMID:2920185  
A:Accession: S02766  
A:Molecule type: mRNA  
A:Residues: 1-376 <EMR>  
A:Cross-references: EMBL:X14778; NID:G57284; PIDN:CAA32085.1; PID:G57285  
C:Superfamily: pulmonary surfactant protein B; saposin repeat homology  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-190/Domain: propeptide #status predicted <PRO>  
F:59-151/Domain: saposin repeat homology <SAP1>  
F:190-277/Domain: saposin repeat homology <SAP2>  
F:191-269/Product: pulmonary surfactant protein B #status predicted <MAT>  
F:266-371/Domain: saposin repeat homology <SAP3>

Query Match 83.1%; Score 343; DB 2; Length 376;  
Best Local Similarity 80.5%; Pred. No. 2, 6e-31;

Matches 62; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Qy 2 PIPPLPYCMLCRALIKRIQAMIPKALRVAAVAVQVCRVPLVAGICQCLAERYSVILDTL 61  
Db 192 PIPPLPYCMLCRALIKRIQAMIPKALRVAAVAVQVCRVPLVAGICQCLAERYSVILDTL 251

Qy 62 LGRMLPOLVCRVLVLRCS 78  
Db 252 LGRVLPOLVCGVLVLRCS 268

RESULT 5  
146531

surfactant protein B - rabbit

C/Species: Oryctolagus cuniculus (domestic rabbit)

C/Date: 14-Feb-1997 #sequence\_revision 14-Feb-1997 #text\_change 20-Aug-1999

C/Accession: I46531

R/Margena, R.K.; Boggaram, V.

Am. J. Physiol. 268, L481-L490, 1995

A/Title: Transcription and mRNA stability regulate developmental and hormonal expression

A/Reference number: I46531; MUID:95208794; PMID:7900830

A/Accession: I46531

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-369 <MAR>

A/Cross-references: EMBL:U17106; NID:9642487; PIDN:AAA67934.1; PID:9642488

A/Genetics:

C/Suprafamily: SP-B

C/Suprafamily: pulmonary surfactant protein B; saposin repeat homology

F:61-153/Domain: saposin repeat homology <SAP1>

Query Match  
Best Local Similarity 80.4%; Score 332; DB 2; Length 369;  
Matches 63; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Qy 1

1 FPIPLPYCMLCRALIKRIQAMIPKALRVAAVAVQVCRVPLVAGICQCLAERYSVILDTL 60

Db 184 FPIPLPYCMLCRALIKRIQAMIPKALRVAAVAVQVCRVPLVAGICQCLAERYSVILDTL 243

Qy 61 LGRMLPOLVCRVLVLRCS 78

Db 244 LGRVLPOLVCGVLVLRCS 261

RESULT 6  
LNG1

pulmonary surfactant protein 9K form - pig

N/Alternate names: low molecular mass surfactant protein type 1

C/Species: Sus scrofa domestica (domestic pig)

C/Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 01-Dec-1995

C/Accession: S00363

R/Curtis, T.; Johnson, J.; Barrois-Soederling, J.; Robertson, B.; Nilsson, G.; Westbe

Eur. J. Biochem. 172, 521-525, 1988

A/Title: Low-molecular-mass surfactant protein type 1. The primary structure of a hydrop

A/Reference number: S00363; MUID:8166729; PMID:3350011

A/Accession: S00363

A/Molecule type: protein

A/Residues: 1-79 <CCR>

C/Comment: Pulmonary surfactant protein is a phospholipid-protein complex, which reduces

C/Suprafamily: pulmonary surfactant protein B; saposin repeat homology

C/Keywords: alveolar proteinosis; gaseous exchange; lipoprotein; lung; pulmonary surfact

F:1-79/Domain: saposin repeat homology <SAP>

Query Match  
Best Local Similarity 77.7%; Score 321; DB 1; Length 79;  
Matches 61; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

Db 61 LLDRTLPOLVCGVLVLRCS 78

RESULT 7

A29667

pulmonary surfactant protein B - bovine (fragment)

N/Alternate names: pulmonary surfactant 1-Pne protein

C/Species: Bos primigenius taurus (cattle)

C/Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 16-Oct-1998

C/Accession: A29667; S02317

R/Olafson, R.W.; Rink, U.; Kiehl, S.; Yu, S.H.; Chung, J.; Harding, P.G.R.; Possmayer

Biochem. Biophys. Res. Commun. 148, 1406-1411, 1987

A/Title: Protein sequence analysis studies on the low molecular weight hydrophobic prote

A/Reference number: A90137; MUID:88077030; PMID:3689402

A/Accession: A29667

A/Molecule type: protein

A/Residues: 1-60 <OLA>

R/Yu, S.H.; Chung, W.; Olafson, R.W.; Harding, P.G.R.; Possmayer, F.

Biochim. Biophys. Acta 921, 437-448, 1987

A/Title: Characterization of the small hydrophobic proteins associated with pulmonary su

A/Reference number: S02317; MUID:88025156; PMID:3663690

A/Accession: S02317

A/Molecule type: protein

A/Residues: 1-10 <YU1>

C/Suprafamily: pulmonary surfactant protein B; saposin repeat homology

Qy 1

1 FPIPLPYCMLCRALIKRIQAMIPKALRVAAVAVQVCRVPLVAGICQCLAERYSVIL 57

Db 1 FPIPLPYCMLCRALIKRIQAMIPKALRVAAVAVQVCRVPLVAGICQCLAERYSVIL 57

RESULT 8

T00207

p109 protein - silkworm

C/Species: Bombyx mori (silkworm)

C/Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 21-Jul-2000

C/Accession: T00207

R/Tambunan, J.; Chang, P.K.; Li, H.; Natori, M.

Gene 212, 287-293, 1998

A/Title: Molecular cloning of a cDNA encoding a silkworm protein which contains the con

A/Reference number: Z14124; MUID:98278844; PMID:9611271

A/Accession: T00207

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-965 <TAM>

A/Cross-references: EMBL:AB008449; NID:92575864; PIDN:BA23126.1; PID:92575865

C/Suprafamily: saposin repeat homology

F:778-870/Domain: saposin repeat homology <SAP3>

Query Match  
Best Local Similarity 22.6%; Score 93.5; DB 2; Length 965;  
Matches 22; Conservative 19; Mismatches 29; Indels 7; Gaps 3;

Qy 8 CMLCRALIKRIQAMIPKALRVAAVAVQVCRVPLVAGICQCLAERYSVILDTL 61  
Db 217 CQICLPMVAVQARQOLQSNRQDEIKVFEQSCILPIKFAEG-CMLADEFVELIETL 275

Qy 62 LGRMLPOLVCRVLVLRCS 78  
Db 276 ASEMFPQAVCSVAGLCN 292

RESULT 9  
S21770

saposin-C - bovine

C/Species: Bos primigenius taurus (cattle)

C/Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 31-Oct-1997

C/Accession: S21770

R/Sano, A.; Mizuno, T.; Kondoh, K.; Hinenno, T.; Ueno, S.; Kakimoto, Y.; Morita, N.

Biochim. Biophys. Acta 1120, 75-80, 1992  
 A>Title: Saposin-C from bovine spleen; complete amino acid sequence and relation between  
 A:Reference number: S21770; MUID:92207994; PMID:1554743  
 A:Accession: S21770  
 A>Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-80 <SAN>  
 C:Superfamily: saposin, saposin repeat homology

Query Match 22.5%; Score 93; DB 2; Length 80;  
 Best Local Similarity 27.6%; Pred. No. 0.0013;  
 Matches 21; Conservative 15; Mismatches 36; Indels 4; Gaps 1;

Qy 7 YCMLRALIKRIQAMIPKQ---ALRVAVQCVPLVAGGICCCLAERYSVILLDTLL 62  
 Db 4 YCQVEFVYKVAKIDNNRTEETLHLDKVCSTLPISLAQCCBEVDYGRSITSLIL 63  
 Qy 63 GRMLPOLVCRLVLRCS 78  
 Db 64 DEASPELVCSMLHCS 79

RESULT 10  
 SAHUP  
 saposin precursor [validated] - human  
 N:Alternate names: cerebroside sulfate activator protein; co-beta-glucosidase; component  
 ein (SAP); sphingolipid activator protein A2; sulfatide sulfatase activator protein  
 N:Contains: prosaposin; saposin A; saposin B; saposin C; saposin D  
 C:Species: Homo sapiens (man)  
 C:Date: 30-Jun-1992 #sequence-revision 17-Nov-1995 #text-change 08-Dec-2000  
 C:Accession: JX0061; A57368; A42003; B42003; C42003; A30367; S34740; S36140; S36  
 0226; I37265; I37264  
 J: Nakano, T.; Sandhoff, K.; Stuenkel, J.; Christomanou, H.; Suzuki, K.  
 J. Biochem. 105, 152-154, 1989  
 A>Title: Structure of full-length cDNA coding for sulfatide activator, a Co-beta-glucosi  
 A:Reference number: JX0061; MUID:89255151; PMID:2498298  
 A:Accession: JX0061  
 A:Molecule type: mRNA  
 A:Residues: 1-527 <NAK>  
 A:Cross-references: GB:D00422; NID:9220063; PIDN:BA00021.1; PID:9220064  
 A>Note: alternative splice form 1  
 A:Accession: A57368  
 A:Molecule type: mRNA  
 A:Residues: 1-259, 263-527 <NA2>  
 A:Cross-references: GB:J03015; GB:J03086; NID:9337755; PIDN:AA059494.1; PID:9337756  
 A>Note: alternative splice form 2  
 R:Rortman, E.G.; Scheinker, V.; Grabowski, G.A.  
 Genomics 13, 312-318, 1992  
 A>Title: Structure and evolution of the human prosaposin chromosomal gene.  
 A:Reference number: A42003; MUID:92307663; PMID:1612590  
 A:Accession: A42003  
 A:Molecule type: DNA  
 A:Residues: 50-140 <ROR>  
 A:Cross-references: GB:M6181  
 A>Note: sequence extracted from NCBI backbone (NCBIN:107235, NCBIP:107236)  
 A:Accession: B42003  
 A:Molecule type: DNA  
 A:Residues: 185-259; 263-276 <RO2>  
 A>Note: sequence extracted from NCBI backbone (NCBIN:107235, NCBIP:107237)  
 A:Accession: C42003  
 A:Molecule type: DNA  
 A:Residues: 305-393 <RO3>  
 A>Note: sequence extracted from NCBI backbone (NCBIN:107235, NCBIP:107238); sequence inc  
 A:Accession: D42003  
 A:Molecule type: DNA  
 A:Residues: 399-487 <RO4>  
 A>Note: sequence extracted from NCBI backbone (NCBIN:107235, NCBIP:107239); sequence inc  
 R:Rortman, E.G.; Grabowski, G.A.  
 Genomics 5, 486-492, 1989  
 A>Title: Molecular cloning of a human co-beta-glucosidase cDNA: evidence that four sphin  
 A:Reference number: A30367; MUID:90129043; PMID:2515150  
 A:Accession: A30367  
 A:Molecule type: mRNA

A:Residues: 1-259, 263-527 <RO5>  
 A:Cross-references: GB:J03077; NID:9183230; PIDN:AA052560.1; PID:9183231  
 A>Note: alternative splice form 2  
 R:Hiraiwa, M.; O'Brien, J.S.; Kishimoto, Y.; Galdzicka, M.; Fluharty, A.L.; Gims, E.I.  
 Arch. Biochem. Biophys. 304, 110-116, 1993  
 A>Title: Isolation, characterization, and proteolysis of human prosaposin, the precurs  
 A:Reference number: S34740; MUID:93311991; PMID:8323276  
 A:Accession: S34740  
 A:Molecule type: protein  
 A:Residues: 17-24; 165-172; 180-189; 301-305 <HR>  
 R:Tyndel, J.; Palmer, D.N.; Baumann, M.; Haltic, M.  
 FBS Lett. 330, 8-12, 1993  
 A>Title: Storage of saposins A and D in infantile neuronal ceroid-lipofuscinosis.  
 A:Reference number: S36140; MUID:93380576; PMID:8370464  
 A:Accession: S36140  
 A:Molecule type: protein  
 A:Residues: 'XX', 62, 'X', 64-65, 'X', 67-79, 'X', 81-84 <TY>  
 A>Note: saposin A  
 A:Accession: S36141  
 A:Molecule type: protein  
 A:Residues: 'XXX', 413-414, 'X', 416-428, 'X', 430-434 <TY2>  
 A>Note: saposin D  
 R:Holtschmidt, H.; Sandhoff, K.; Kwon, H.Y.; Haezler, K.; Nakano, T.; Suzuki, K.  
 J. Biol. Chem. 266, 7556-7560, 1991  
 A>Title: Sulfatide activator protein. Alternative splicing that generates three mRNAs a  
 A:Reference number: S36988; MUID:91210267; PMID:2019586  
 A:Accession: S36988  
 A>Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-240, 'S', 242-527 <HOL>  
 A:Cross-references: EMBL:M60255; NID:9337759; PIDN:AA06594.1; PID:9337760  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1991  
 A:Accession: S36989  
 A>Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-240, 'S', 242-259, 263-527 <HO2>  
 A:Cross-references: EMBL:M60257; NID:9337764; PIDN:AA06595.1; PID:9337765  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1991  
 A:Accession: S36990  
 A>Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-240, 'S', 242-259, 261-527 <HO3>  
 A:Cross-references: EMBL:M60258; NID:9337766; PIDN:AA06596.1; PID:9337767  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1991  
 A:Accession: S36991  
 A>Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-240, 'S', 242-259, 261-527 <HO3>  
 A:Cross-references: EMBL:M60258; NID:9337766; PIDN:AA06596.1; PID:9337767  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1991  
 R:Kondo, K.; Hienno, T.; Sano, A.; Kakimoto, Y.  
 Biochem. Biophys. Res. Commun. 181, 286-292, 1991  
 A>Title: Isolation and characterization of prosaposin from human milk.  
 A:Reference number: P50330; MUID:92068206; PMID:1958198  
 A:Accession: P50330  
 A:Molecule type: protein  
 A:Residues: 17-24, 'X', 26 <KON>  
 A:Experimental source: milk  
 R:Kretz, K.A.; Carson, G.S.; Morimoto, S.; Kishimoto, Y.; Fluharty, A.L.; O'Brien, J.S.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 2541-2544, 1990  
 A>Title: Characterization of a mutation in a family with saposin B deficiency: a glycos  
 A:Reference number: A35985; MUID:90207231; PMID:2320574  
 A:Accession: A35985  
 A:Molecule type: mRNA  
 A:Residues: 213-221 <KRE>  
 A:Cross-references: GB:M52221  
 A:Accession: B35986  
 A>Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-259, 263-527 <KR2>  
 A:Cross-references: GB:M52221; NID:9337761; PIDN:AA060303.1; PID:9337762  
 A:Experimental source: lymphoblast  
 A:Accession: C35985  
 A:Molecule type: mRNA  
 A:Residues: 213-216, 'T', 218-221 <KR3>  
 A>Note: sequence from patients with activator-deficient metachromatic leukodystrophy; t

R.Fuerst, W.; Schubert, J.; Machleidt, W.; Meyer, H.E.; Sandhoff, K.  
 Eur. J. Biochem. 192, 709-714, 1990  
 A>Title: The complete amino-acid sequences of human ganglioside GM2 activator protein and  
 A:Reference number: S13195; MUID:91006165; PMID:2209618  
 A:Accession: S13196  
 A:Molecule type: protein  
 A:Residues: 195-259,263-277 <FUE>  
 R.Morimoto, S.; Martin, B.M.; Yamamoto, Y.; Kretz, K.A.; O'Brien, J.S.; Kishimoto, Y.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 3389-3393, 1989  
 A>Title: Saposin A: second cerebroside activator protein.  
 A:Reference number: A32784; MUID:89240739; PMID:2717620  
 A:Accession: A32784  
 A:Molecule type: protein  
 A:Residues: 60-84,86-107,109-119,125-134 <MOR>  
 R.O'Brien, J.S.; Kretz, K.A.; Dewji, N.; Wenger, D.A.; Beech, F.; Fluharty, A.L.  
 Science 241, 1098-1101, 1988  
 A>Title: Coding of two sphingolipid activator proteins (SAP-1 and SAP-2) by same genetic  
 A:Reference number: A41240; MUID:88321660; PMID:2842863  
 A:Accession: A41240  
 A:Molecule type: mRNA  
 A:Residues: 18-259,263-299, 'D', 301-302, 'D', 304-527 <ONB>  
 R.Dewji, N.N.; Wenger, D.A.; O'Brien, J.S.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 8652-8656, 1987  
 A>Title: Nucleotide sequence of cloned cDNA for human sphingolipid activator protein 1 p  
 A:Reference number: S02289; MUID:88068647; PMID:2825202  
 A:Accession: S02289  
 A>Status: significant sequence differences  
 A:Molecule type: mRNA  
 A:Cross-references: EMBL:J03015  
 A>Note: this sequence corrected by A41240  
 R:Kleinhardt, T.; Christomanu, H.; Brannitzler, G.  
 Biol. Chem. Hoppe-Seyler 369, 1161-1165, 1988  
 A>Title: Complete amino-acid sequence of the naturally occurring A(2) activator protein  
 A:Reference number: S02028; MUID:89207116; PMID:3242355  
 A:Accession: S02028  
 A:Molecule type: protein  
 A:Residues: 195-259,263-276 <KLE>  
 R.Fuerst, W.; Machleidt, W.; Sandhoff, K.  
 Biol. Chem. Hoppe-Seyler 369, 317-328, 1988  
 A>Title: The precursor of sulfatide activator protein is processed to three different pr  
 A:Reference number: S00813; MUID:89000190; PMID:3048308  
 A:Accession: S00813  
 A:Molecule type: protein  
 A:Residues: 410-487 <FU2>  
 R.Kleinhardt, T.; Christomanu, H.; Brannitzler, G.  
 Biol. Chem. Hoppe-Seyler 368, 1571-1578, 1987  
 A>Title: Complete amino-acid sequence and carbohydrate content of the naturally occurring  
 A:Reference number: S00226; MUID:88163077; PMID:3442600  
 A:Accession: S00226  
 A:Molecule type: protein  
 A:Residues: 314-393 <KLT>  
 R.Vaccaro, A.M.; Salvatelli, R.; Barca, A.; Tatti, M.; Claffoni, F.; Maras, B.; Sticiliano,  
 U. Biol. Chem. 270, 9951-9960, 1995  
 A>Title: Structural analysis of saposin C and B. Complete localization of disulfide brid  
 A:Reference number: A57297; MUID:95247790; PMID:7730378  
 A:Accession: A57297  
 A:Content: annotation: disulfide bonds: glycosylation  
 R.Holtschmidt, H.; Sandhoff, K.; Fuerst, W.; Kwon, H.Y.; Schnabel, D.; Suzuki, K.  
 FEBS Lett. 280, 267-270, 1991  
 A>Title: The organization of the gene for the human cerebroside sulfate activator protei  
 A:Reference number: I37264; MUID:91192146; PMID:2013321  
 A:Accession: I37265  
 A>Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 59-125 <RES>  
 A:Cross-references: EMBL:X57107; NID:G30234; PIDN:CAA40391.1; PID:G30235  
 A:Accession: I37264  
 A>Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 307-516 <RE2>  
 A:Cross-references: EMBL:X57108; NID:G30232; PIDN:CAA40392.1; PID:G1565257  
 A>Note: sequence revised relative to PID:G30233 (corrected coding region)

C:Genetics:  
 A:Gene: GDB:PSAP; GLBA  
 A:Cross-references: GDB:120366; OMIM:176801  
 A:Map position: 10q22.1-10q22.1  
 A:Introns: 83/3; 338/3; 401/1; 453/3; 480/3  
 A>Note: defects in this gene may cause variant Gaucher disease, variant Tay-Sachs disea  
 A>Note: list of introns is incomplete

Query Match 22.0%; Score 91; DB 1; Length 527;  
 Best Local Similarity 27.6%; Pred. No. 0.013;  
 Matches 21; Conservative 15; Mismatches 36; Indels 4; Gaps 1;

Oy 7 YCWLGRALIKRIQAMIPKGA---LRVAVAQVCRVPLVAGIGCQCLAERYSVILDTLL 62  
 Db 317 YCEVCEFLVEVTKTILNMNTEKEILAFPMCKSKLRSLEQEVVDYTGSSILSL 376

Oy 63 GRMLPOLVCRVLVRCSS 78  
 Db 377 EEVSPFLVCSMLHLCS 392

RESULT 11  
 A28716  
 N:Alternate names: cerebroside sulfate activator protein; co-beta-glucosidase; component  
 ein (SAP); sphingolipid activator protein A2; sulfated glycoprotein 1; sulfatide sulfat  
 N:Contents: prosaposin; saposin A; saposin B; saposin C; saposin D  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 10-Sep-1999 #sequence (Norway rat)  
 C:Accession: A28716  
 R:Collard, M.W.; Sylvester, S.R.; Teunisse, J.K.; Griswold, M.D.  
 Biochemistry 27, 4557-4564, 1988  
 A>Title: Biosynthesis and molecular cloning of sulfated glycoprotein 1 secreted by rat

A:Reference number: A28716; MUID:89000647; PMID:3048385  
 A:Accession: A28716  
 A:Molecule type: mRNA  
 A:Cross-references: GB:M19936; NID:G206904; PIDN:AAA42136.1; PID:G206905  
 A>Note: parts of this sequence, including the amino end of the mature protein, were det  
 C:Function:

A:Description: saposins bind sphingolipids, form hydrophilic complexes and make them ac  
 A:Pathway: sphingolipid catabolism  
 A>Note: saposins A and C (SAP-2) activate hydrolysis of glucocerebroside by beta-glucos  
 A>Note: saposin B (SAP-1) activates hydrolysis of galactocerebroside sulfate by arylsul  
 A>Note: saposin D activates hydrolysis of sphingomyelin by sphingomyelin phosphodiester  
 C:Superfamily: saposin; saposin repeat homology  
 C:Keywords: alternative splicing; glycoprotein; lysosomal storage disease; lysosome; spi  
 F:1-16/Domain: signal sequence #status predicted <SIG>  
 F:15-54/Domain: prosaposin #status predicted <PRO>  
 F:55-148/Domain: saposin repeat homology <SAP1>  
 F:60-143/Domain: saposin A #status predicted <SAP1>  
 F:189-280/Domain: saposin repeat homology <SAP2>  
 F:194-273/Domain: saposin B #status predicted <SAB1>  
 F:306-397/Domain: saposin repeat homology <SAP3>  
 F:310-389/Domain: saposin C #status predicted <SAPC>  
 F:431-522/Domain: saposin repeat homology <SAP4>  
 F:437-514/Domain: saposin D #status predicted <SAPD>  
 F:63-118,66-132,94-106,439-512,442-506,470-481/Disulfide bonds: #status predicted  
 F:80,214,331,456/Binding site: carbohydrate (Aan) (covalent) #status predicted  
 F:197-270,200-264,229-240,314-387,317-381,345-356/Disulfide bonds: #status predicted

Query Match 21.1%; Score 87; DB 1; Length 554;  
 Best Local Similarity 23.7%; Pred. No. 0.04;  
 Matches 18; Conservative 21; Mismatches 33; Indels 4; Gaps 1;

Oy 7 YCWLGRALIKRIQAMIPKGA---LRVAVAQVCRVPLVAGIGCQCLAERYSVILDTLL 62  
 Db 313 FCQVQCLVWRKLSLILNNATELLIKGSKKCSLLPAPASTKCGEVIVTFGSLDLVLM 372

Oy 63 GRMLPOLVCRVLVRCSS 78  
 Db 373 HEVSPFLVCSMLHLCS 398



RESULT 12  
A32026  
glucosylceramide beta-glucosidase activator protein SAP-2 - guinea pig  
C:Species: Cavia porcellus (guinea pig)  
C:Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 17-May-1996  
C:Accession: A32026  
R:Sano, A.; Radin, N.S.; Johnson, L.L.; Tarr, G.E.  
J. Biol. Chem. 263, 19597-19601, 1988  
A>Title: The activator protein for glucosylceramide beta-glucosidase from guinea pig liv  
A:Reference number: A32026; MUID:89066787; PMID:3198642  
A:Accession: A32026  
A:Molecule type: protein  
A:Residues: 1-81 <SAP>  
C:Superfamily: saposin; saposin repeat homology  
F:1-81/Domain: saposin repeat homology <SAP>

Query Match 20.1%; Score 83; DB 2; Length 81;  
Best Local Similarity 26.5%; Pred. No. 0.018;  
Matches 18; Conservative 14; Mismatches 32; Indels 4; Gaps 1;

Qy 8 CMLCRALIKRIQAMIPKGLARV---AVAVQCRVPLVAGICQCLAEYSVILDTLLG 63  
Db 5 CKACQYVVKWKLIDNNRTEKIHLDVSCALPESVSEVQEVVDYGDYSVALLLQ 64  
Qy 64 RMLPOLVC 71  
Db 65 EWSPELVC 72

RESULT 13  
S59823  
probable membrane protein YPR164w - Yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein P9325.4  
C:Species: Saccharomyces cerevisiae  
C:Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 24-Nov-1999  
C:Accession: S59823  
R:Hallsworth, K.  
submitted to the EMBL Data Library, April 1995  
A:Description: The sequence of S. cerevisiae cosmid 9325.  
A:Reference number: S59821.  
A:Accession: S59823  
A:Molecule type: DNA  
A:Residues: 1-1407 <HAL>  
A:Cross-references: EMBL:U25840; NID:g786286; PID:g786289; MIPS:YPR164w  
C:Genetics:  
A:Gene: SGD:KIM3  
A:Cross-references: SGD:S0006368; MIPS:YPR164w  
A:Map position: 16R  
C:Superfamily: Saccharomyces cerevisiae probable membrane protein YPR164w  
C:Keywords: transmembrane protein  
F:218-234/Domain: transmembrane #status predicted <TM1>  
F:320-336/Domain: transmembrane #status predicted <TM2>  
F:488-504/Domain: transmembrane #status predicted <TM3>  
F:870-886/Domain: transmembrane #status predicted <TM4>  
F:1019-1035/Domain: transmembrane #status predicted <TM5>

Query Match 15.4%; Score 63.5; DB 2; Length 1407;  
Best Local Similarity 26.8%; Pred. No. 45;  
Matches 15; Conservative 11; Mismatches 19; Indels 11; Gaps 2;

Qy 1 PPIPLPYCMLCRALIKRIQAMIPKGLRVAVAVQCRVPLVAGICQCLAEYSV 56  
Db 408 FPAPL-----LLKLQELHPHTFKF---QYCISSSGNICFCVTERSTIV 452

RESULT 14  
YGCVC  
alpha-aminoadipyl-cysteinyI-valine synthetase (EC 6.-.-.-) - fungus (Acremonium chrysog  
N:Alternate names: ACV synthetase  
C:Species: Acremonium chrysogenum  
C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 03-Nov-2000  
C:Accession: A38531

R:Gutierrez, S.; Diez, B.; Montenegro, E.; Martin, J.F.  
J. Bacteriol. 173, 2354-2365, 1991  
A>Title: Characterization of the Cephalosporin biosynthetic genes and evidence of multiple functional doma  
ster of early cephalosporin biosynthetic genes  
A:Reference number: A38531; MUID:91177827; PMID:1706706  
A:Accession: A38531  
A:Molecule type: DNA  
A:Residues: 1-3712 <GUT>  
C:Comment: This multifunctional enzyme catalyzes the formation of alpha-aminoadipyl-cys  
tryng out the polymerization steps to form the tripeptide, which is the first common in

C:Genetics:  
A:Gene: pcbAB  
C:Superfamily: alpha-aminoadipyl-cysteinyI-valine synthetase; acetate-CoA ligase homolo  
C:Keywords: carrier protein; cephalosporin biosynthesis; duplication; ligase; penicillin  
F:306-774/Domain: acetate-CoA ligase homology <ACLI>  
F:793-863/Domain: acyl carrier protein homology <ACPI>  
F:1408-1866/Domain: acetate-CoA ligase homology <ACLI2>  
F:11882-1952/Domain: acyl carrier protein homology <ACPI2>  
F:2482-2942/Domain: acetate-CoA ligase homology <ACLI3>  
F:2958-3026/Domain: acyl carrier protein homology <ACPI3>  
F:827,1916,2990/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 15.4%; Score 63.5; DB 1; Length 3712;  
Best Local Similarity 37.2%; Pred. No. 1.1e-02;  
Matches 16; Conservative 6; Mismatches 20; Indels 1; Gaps 1;

Qy 27 LRAVAVQCRVPLVAGICQCLAEYSVILDTLLGRMLPOL 69  
Db 3237 LRKTVAGSFETPLICMGVLOC-QEKSRETTALLSKACPAL 3278

RESULT 15  
S44252  
raffinose operon regulatory protein rafr - Pedicoccus pentosaceus  
C:Species: Pedicoccus pentosaceus  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 05-Nov-1999  
C:Accession: S44252  
R:leenhout, K.K.J.; Bolhuis, A.A.; Kók, J.J.; Venema, G.G.  
submitted to the EMBL Data Library, April 1994  
A:Description: The sucrose and raffinose operons of Pedicoccus pentosaceus PPE1.0.  
A:Reference number: S44252  
A:Accession: S44252  
A:Molecule type: DNA  
A:Residues: 1-277 <LEE>  
A:Cross-references: EMBL:Z32771; NID:g493728; PIDN:CA83663.1; PID:g475963  
C:Genetics:  
A:Gene: rafr  
C:Keywords: DNA binding; transcription regulation

Query Match 15.3%; Score 63; DB 2; Length 277;  
Best Local Similarity 31.9%; Pred. No. 11;  
Matches 23; Conservative 14; Mismatches 25; Indels 10; Gaps 3;

Qy 4 PIPYCMCLCRALIKRIQAMIPKGLRVAVAVQCRVPLVAGICQCLAEYSV-----IL 57  
Db 85 PKWYFWMIGSAGIRIEAML-SGSL--LAOKCYLRVQNGHLYADLSLVKLVHPNSLI 140  
Qy 58 IDTLIGRMIPOL 69  
Db 141 NDVLGSLIYRL 152

Search completed: December 3, 2003, 15:53:27  
Job time : 24.1088 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 15:43:05 ; Search time 13.9728 Seconds

(without alignments)  
265.882 Million cell updates/sec

Title: US-09-788-308D-2

Sequence: 1 FPIPLPYCWLKRALIKRIQA.....TLIGRLPLQVLCRLVLRCSM 79

Scoring table: BLASTSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	407	98.5	381	1	PSPB_HUMAN
2	350	84.7	363	1	PSPB_CANPA
3	344	83.3	370	1	PSPB_RABIT
4	343	83.1	376	1	PSPB_RAT
5	343	83.1	377	1	PSPB_MOUSE
6	321	77.7	79	1	PSPB_PIG
7	267	64.6	79	1	PSPB_BOVIN
8	92.5	22.4	525	1	SAP_BOVIN
9	91	22.0	524	1	SAP_HUMAN
10	87	21.1	554	1	SAP_RAT
11	83	20.1	81	1	SAP_CAVPO
12	80.5	19.5	518	1	SAP_CHICK
13	66	16.0	318	1	FX24_HUMAN
14	63.5	15.4	3712	1	ACVS_CEPAC
15	63	15.3	277	1	RAFR_PEDP
16	63	15.3	557	1	SAP_MOUSE
17	62.5	15.1	92	1	YC23_METUA
18	62.5	15.1	457	1	VIPR_HUMAN
19	61.5	14.9	277	1	J11_HCMYA
20	61.5	14.9	469	1	GATB_THETH
21	60.5	14.6	608	1	UL27_HCMYA
22	59.5	14.4	310	1	O2A4_HUMAN
23	59.5	14.4	417	1	PROA_ECO57
24	59.5	14.4	417	1	PROA_ECOLI
25	59.5	14.4	446	1	NU4M_DROYA
26	59	14.3	83	1	YG41_STRAS
27	59	14.3	407	1	NGRE_YERPE
28	59	14.3	622	1	FACG_HUMAN
29	58.5	14.2	651	1	PRIA_MYCLE
30	58.5	14.2	129	1	NKL_PIG
31	58.5	14.2	295	1	RHLA_PSRSE
32	58.5	14.2	446	1	NU4M_DROME
33	58	14.0	293	1	IF35_ARATH

## ALIGNMENTS

34	58	14.0	535	1	GRB7_MOUSE
35	58	14.0	759	1	NAHE_ONCMY
36	57.5	13.9	359	1	PTWC_ECOLI
37	57.5	13.9	625	1	PTBA_ECOLI
38	57	13.8	396	1	SPOB_MOUSE
39	57	13.8	740	1	GNT5_RAT
40	56.5	13.7	122	1	FOLB_CHLMU
41	56.5	13.7	145	1	GNLY_HUMAN
42	56.5	13.7	297	1	PANE_LISMO
43	56.5	13.7	430	1	CIMC_HUMAN
44	56.5	13.7	430	1	CIMC_RAT
45	56	13.6	237	1	LMPI_CAEEL

## RESULT 1

ID	PSPB_HUMAN	STANDARD	PRT	381 AA
AC	P07988; O96R04;			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Pulmonary surfactant-associated protein B precursor (SP-B) (6 kDa protein) (Pulmonary surfactant-associated proteolipid SPu(Phe)) (18 kDa pulmonary-surfactant protein)			
DE	kDa pulmonary-surfactant protein			
GN	SPBP OR SPBP3.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxId=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=89170128; PubMed=2924687;			
RA	Pilot-Matias T.J., Kistler S.E., Fox J.L., Kropp K., Glaeser S.W.,			
RA	Whitsett J.A.;			
RT	"Structure and organization of the gene encoding human pulmonary			
RT	surfactant proteolipid SP-B.";			
RL	DNA 8:75-86(1985).			
RN	[2]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 201-214.			
RC	TISSUE=Lung;			
RX	MEDLINE=87250653; PubMed=3597440;			
RA	Jacob K.A., Phelps D.S., Steinbrink R., Fisch J., Kriz R.,			
RA	Mitsch U., Dougherty J.P., Taesch H.W., Flores J.;			
RT	"Isolation of a cDNA clone encoding a high molecular weight precursor			
RT	to a 6-kDa pulmonary surfactant-associated protein.";			
RL	J. Biol. Chem. 262:9808-9811(1987).			
RN	[3]			
RP	SEQUENCE FROM N.A., AND VARIANTS 116-131; PHE-176 AND HIS-272.			
RA	Rieder M.J., Carrington D.P., Chung M.-W., Lee K.L., Poel C.L., Yi Q.,			
RA	Nickerson D.A.;			
RT	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE OF 99-381 FROM N.A.			
RX	MEDLINE=87231940; PubMed=3035561;			
RA	Glaeser S.W., Korthagen T.R., Weaver T., Pilot-Matias T., Fox J.L.,			
RA	Whitsett J.A.;			
RT	"cDNA and deduced amino acid sequence of human pulmonary surfactant-			
RT	associated proteolipid SPu(Phe).";			
RL	Proc. Natl. Acad. Sci. U.S.A. 84:4007-4011(1987).			
RN	[5]			
RP	SEQUENCE OF 139-381 FROM N.A.			
RX	MEDLINE=8819786; PubMed=3343343;			
RA	Revak S.D., Merritt T.A., Degryse E., Stefani L., Courtney M.,			
RA	Hallman M., Cochran C.G.;			
RT	"Use of human surfactant low molecular weight apoproteins in the			
RT	reconstitution of surfactant biologic activity.";			
RL	J. Clin. Invest. 81:826-833(1988).			
RN	[6]			
RP	SEQUENCE OF 201-279, AND DISULFIDE BONDS.			
RX	MEDLINE=92233937; PubMed=1568474;			

RA Johanson J., Joernvall H., Cursstedt T.;  
 RT "Human surfactant polypeptide SP-B. Disulfide bridges, C-terminal  
 RL end, and peptide analysis of the airway form.";  
 RN J. Biol. Chem. 268:165-167 (1992).  
 RP STRUCTURE BY FTIR OF 201-225.  
 RX MEDLINE=20256237; PubMed=10798379;  
 RA Gordon L.M., Lee K.Y., Lipp M.M., Zasadzinski J.A., Maltzer F.J.,  
 RA Sherman M.A., Waring A.J.;  
 RT "Conformational mapping of the N-terminal segment of surfactant  
 RT protein B in lipid using 13C-enhanced Fourier transform infrared  
 RT spectroscopy.";  
 RL J. Pept. Res. 55:330-347 (2000).  
 RN [8].  
 RP VARIANT ILE-131.  
 RX MEDLINE=20525299; PubMed=11076040;  
 RA Lin Z., Pearson C., Chinchilli V., Pleetschmann S.M., Luo J., Pison U.,  
 RA Floros J.;  
 RT "Polymorphisms of human SP-A, SP-B, and SP-D genes: association of  
 RT SP-B Thr311le with ARDS.";  
 RL Clin. Genet. 58:181-191 (2000).  
 CC -1- FUNCTION: PULMONARY SURFACTANT ASSOCIATED PROTEINS PROMOTE  
 CC ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-  
 CC LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES. SP-B INCREASES  
 CC THE COLLAPSE PRESSURE OF PALMITIC ACID TO NEARLY 70 MILLINEWTONS  
 CC PER METER.  
 CC -1- SUBUNIT: Homodimer; disulfide-linked.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- POLYMORPHISM: There seems to be an association between the Thr-  
 CC 131-Ile polymorphism and acute respiratory distress syndrome  
 CC (ARDS).  
 CC -1- DISEASE: Defects in SFTPB are the cause of congenital pulmonary  
 CC alveolar proteinosis; an autosomal recessive fatal respiratory  
 CC disease.  
 CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%  
 CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,  
 CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL  
 CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).  
 CC -1- SIMILARITY: Contains 1 saposin A-type domain.  
 CC -1- SIMILARITY: Contains 3 saposin B-type domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; M24461; AAB59541.1; -;  
 DR EMBL; J02761; AAB60212.1; -;  
 DR EMBL; AF400074; AAK77913.1; -;  
 DR EMBL; M16764; AAB88099.1; ALT\_INIT.  
 DR EMBL; M19097; AAB36628.1; -;  
 DR PIR; A31361; LNHUB.  
 DR PDB; 1DFW; 24-OCT-01.  
 DR PDB; 1KMR; 07-AUG-02.  
 DR GeneW; HGNC:10801; SFTPB.  
 DR MIM; 178640; -;  
 DR MIM; 265120; -;  
 DR GO; GO:0007397; P.histogenesis and organogenesis; TAS.  
 DR GO; GO:0007585; P.respiratory gaseous exchange; TAS.  
 DR InterPro; IPR003119; SAPB.  
 DR InterPro; IPR000004; SAPB.  
 DR InterPro; IPR003258; Surfactant\_B.  
 DR Pfam; PF02199; SAPB\_1.  
 DR Pfam; PF03184; SAPB\_1; 1.  
 DR Pfam; PF03489; SAPB\_2; 3.  
 DR Pfam; PD001732; SAPB sub; 1.  
 DR ProDom; PD008002; Surfactant\_B; 1.  
 DR SMART; SM00162; SAPA; 1.  
 DR SMART; SM00118; SAPB; 3.  
 KW Surface film; Gaseous exchange; Glycoprotein; Repeat; Polymorphism;

KW 3D-structure. 1 200  
 FT PROPEP 201 279  
 FT CHAIN 201 279  
 FT  
 FT PROPEP 280 381  
 FT DOMAIN 28 61  
 FT DISULFID 208 277  
 FT DISULFID 211 271  
 FT DISULFID 235 246  
 FT DISULFID 248 248  
 FT CARBOHYD 129 129  
 FT CARBOHYD 311 311  
 FT VARIANT 131 131  
 FT VARIANT 176 176  
 FT VARIANT 176 176  
 FT VARIANT 228 228  
 FT VARIANT 228 228  
 FT VARIANT 228 228  
 FT VARIANT 272 272  
 FT VARIANT 272 272  
 FT CONFLICT 178 178  
 FT CONFLICT 318 318  
 FT HELIX 208 221  
 SQ SEQUENCE 361 AA; 42117 MW; 9FD7F6678A35153 CRC64;  
 Query Match 98.5%; Score 407; DB 1; Length 361;  
 Best Local Similarity 98.7%; Pred. No. 3.7e-40;  
 Matches 78; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 PPIPLPYCMLCRALIKRIQMIKGALEVAVAQVRVPLVAGGICCCLAERYSVILLDT 60  
 DB 201 PPIPLPYCMLCRALIKRIQMIKGALEVAVAQVRVPLVAGGICCCLAERYSVILLDT 260  
 QY 61 LIGRLPOLVCRVLVRCSM 79  
 DB 261 LIGRLPOLVCRVLVRCSM 279  
 RESULT 2  
 PSPB CANFA STANDARD; PRT; 363 AA.  
 ID PSPB CANFA PRT; 363 AA.  
 AC P17129;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Pulmonary surfactant-associated protein B precursor (SP-B) (6 kDa  
 DE protein) (Pulmonary surfactant-associated proteolipid SPL(Phe))  
 DE (Pulmonary surfactant protein 18) (SP 18) (fragment).  
 GN SFTPB OR SFTPB.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Flesipedidae; Canidae; Canis.  
 OC NCBI\_Taxid=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 182-211.  
 RC TISSUE=Lung;  
 RX MEDLINE=87092398; PubMed=3467361;  
 RA Hawgood S., Benson B.J., Schilling J., Damm D., Clements J.A.,  
 RA White R.T.;  
 RT "Nucleotide and amino acid sequences of pulmonary surfactant protein  
 RT SP 18 and evidence for cooperation between SP 18 and SP 28-36 in  
 RT surfactant lipid adsorption.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:66-70 (1987).  
 CC -1- FUNCTION: PULMONARY SURFACTANT-ASSOCIATED PROTEINS PROMOTE  
 CC ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-  
 CC LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES. SP-B INCREASES  
 CC THE COLLAPSE PRESSURE OF PALMITIC ACID TO NEARLY 70 MILLINEWTONS  
 CC PER METER.  
 CC -1- SUBUNIT: Homodimer; disulfide-linked.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%

CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,  
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL  
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).  
CC -1- SIMILARITY: Contains 1 saposin A-type domain.  
CC -1- SIMILARITY: Contains 3 saposin B-type domains.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR	EMBL; M15170; AAA30893.1;	-
DR	PIR; B29072; A29072.	
DR	HSSP; P07988; IDFM.	
DR	InterPro; IPR003119; SAPA.	
DR	InterPro; IPR000004; SApB.	
DR	Pfam; PF02199; SAPA; 1.	
DR	Pfam; PF05184; SApB_1; 1.	
DR	Pfam; PF03489; SApB_2; 3.	
DR	ProDom; PD001732; SApB_sub; 1.	
DR	ProDom; PD008002; Surfactant_B; 1.	
DR	SMART; SM00162; SAPA; 1.	
DR	SMART; SM00118; SApB; 3.	
KW	Surface film; Gaseous exchange; Glycoprotein; Repeat.	
FT	NON TER	1
FT	PROPEP	<1 180
FT	CHAIN	181 259
FT		
FT	PROPEP	260 363
FT	DOMAIN	21 54
FT	DISULFID	188 257
FT	DISULFID	191 251
FT	DISULFID	215 226
FT	DISULFID	228
FT	CARBOHYD	293
FT	CARBONYL	293
QO	SEQUENCE	363 AA; 40180 MW; FADADE02DBE5719 CAC64; N-LINKED (GLYCNC... ) (POTENTIAL).

Query Match	84.7%	Score 350,	DB 1,	length 363;
Best Local Similarity	84.4%	Pred. No. 1.6e-33;		
Matches 65;	Conservative 2;	Mismatches 10;	Indels 0;	Gaps 0;

QY 2 PIPLEPCMLCRALIKRIQAMPKGAIRVAADVCRLVPLVAGGICQCLAEIRYSILLDTL 61  
| | | | | | | | | | | | | | | | : | | |  
DB 182 PIPLEPCMLCRTLIKRIQAMPKGVLAATVGQCCHVPRLVWGICQCLGERYTVDLLDAL 241

QY	62	LGRLPQLVCR	78
Db	242	LGRLPQLVCG	258

RESULT 3  
PSPB\_RABIT  
ID\_PSPB\_RABIT STANDARD; PRT; 370 AA

DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-OCT-1996 (Rel. 34, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Pulmonary surfactant-associated protein B precursor (Sp-B) (6 kDa  
 DE protein) (Pulmonary surfactant-associated proteolipid SPL(Ph)).  
 GN SFPB OR SFPB3.  
 OS *Oryctolagus cuniculus* (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxId=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=89228033; PubMed=2469419;  
 RA Xu J., Richardson C., Ford C., Spencer T., Li-Juan Y., Mackie G.,

RA	Hammond G., Possmayer F.,	"Isolation and characterization of the cDNA for pulmonary surfactant-associated protein-B (SP-B) in the rabbit."	RT
RL	Biochem. Biophys. Res. Commun.	160:325-332(1989).	RL
RN	(2)		RN
RC	SEQUENCE FROM N.A.		RC
RP	STRAIN=New Zealand white;		RP
RX	MEDLINE=95208794; PubMed=7900830;		RX
RA	Margana R.K., Boggaram V.;	"transcription and mRNA stability regulate developmental and hormonal expression of rabbit surfactant protein B gene."	RT
RL	Am. J. Physiol.	268:L481-L490(1995).	RL
RN	(3)		RN
RP	SEQUENCE OF 1-34 FROM N.A.		RP
RX	MEDLINE=96096536; PubMed=8522191;		RX
RA	Lutz P., Anceaschi M., Strayer D.S.;	"The upstream region of the Sp-B gene: intrinsic promoter activity and glucocorticoid responsiveness related to a new DNA-binding protein."	RT
RL	Gene	165:285-290(1995).	RL
CC	-1 FUNCTION: PULMONARY SURFACTANT ASSOCIATED PROTEINS PROMOTE ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES. SP-B INCREASES THE COLLAPSE PRESSURE OF PALMITIC ACID TO NEARLY 70 MILLINEWTONS PER METER.		CC
CC	-1 SUBUNIT: Homodimer; disulfide-linked.		CC
CC	-1 SUBCELLULAR LOCATION: Extracellular.		CC
CC	-1 MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10% PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS, CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL HYDROPHOBIC PROTEINS (SP-B AND SP-C).		CC
CC	-1 SIMILARITY: Contains 1 saposin A-type domain.		CC
CC	-1 SIMILARITY: Contains 3 saposin B-type domain.		CC
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		CC
CC	EMBL; M24901; AAA31466.1; -		CC
DR	EMBL; U17106; AA67934.1; -		DR
DR	EMBL; S80649; AAD1435.1; -		DR
DR	PIR; A32421; LNRBB.		DR
DR	PIR; I46531; I46531.		DR
DR	HSSP; P07988; IDPW.		DR
DR	InterPro; IPR003119; Sapa.		DR
DR	InterPro; IPR000004; SApB.		DR
DR	InterPro; IPR003258; Surfactant_B.		DR
DR	Pfam; PF02199; SAPA; 1.		DR
DR	Pfam; PF05184; SApB_1; 1.		DR
DR	Pfam; PF03489; SApB_2; 3.		DR
DR	ProDom; PD001713; SApB_sub; 1.		DR
DR	ProDom; PD008002; Surfactant_B; 1.		DR
DR	SMART; SM00162; SAPA; 1.		DR
DR	SMART; SM00118; SApB; 3.		DR
KW	Surface film; Gaseous exchange; Glycoprotein; Repeat.		KW
FT	PROPEP	1	FT
FT	CHAIN	185	FT
FT		263	FT
FT			FT
FT	PROPEP	264	FT
FT	DOMAIN	29	FT
FT	DISULFID	192	FT
FT	DISULFID	195	FT
FT	DISULFID	219	FT
FT	DISULFID	232	FT
FT	CARBOHYD	300	FT
FT	CONFLICT	15	FT
FT	CONFLICT	184	FT
FT	CONFLICT	232	FT
FT	CONFLICT	329	FT
FT	CONFLICT	355	FT
FT	ELUTPOLSLTSLSGMDARALCALQAGC -> AAHAPAEPA		FT

FT PROEP 1 190  
FT CHAIN 191 269  
FT PROPEP 270 376  
FT DOMAIN 26 59  
FT DISULFID 198 267  
FT DISULFID 201 261  
FT DISULFID 225 236  
FT DISULFID 238 238  
FT CARBOHYD 306 306  
SQ SEQUENCE 370 AA; 40610 MM; 423047A69B12DCB5 CRC64;  
Query Match 83.1%; Score 344; DB 1; Length 370;  
Best Local Similarity 82.1%; Pred. No. 8e-33;  
Matches 64; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
OY 1 PIPPLPYCMLCRALIKRIQAMIPKCALVAVAAQVGVVPLVAGGICQCLAEERYVILDTL 60  
DB 185 PIPPLPYCMLCRALIKRIQAMIPKCALVAVAAQVGVVPLVAGGICQCLAEERYVILDTL 244  
OY 61 LGRLPOLVCRVLRCS 78  
DB 245 LGRLPOLVCRVLRCS 262  
RESULT 4  
PSPB\_RAT STANDARD; PRT; 376 AA.  
AC P22355;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Pulmonary surfactant-associated protein B precursor (SP-B) (Pulmonary  
DE surfactant-associated proteolipid SPL (Phe)).  
GN SFTPB OR SFTP3.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxId=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89150284; PubMed=2920185;  
RA Emie P.A., Shannon J.M., Mason R.J., Fisher J.H.;  
RT "cDNA and deduced amino acid sequence for the rat hydrophobic  
RT pulmonary surfactant-associated protein, SP-B.";  
RL Biochim. Biophys. Acta 994:215-221(1989).  
CC -1- FUNCTION: PULMONARY SURFACTANT-ASSOCIATED PROTEINS PROMOTE  
CC ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-  
CC LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES. SP-B INCREASES  
CC THE COLLAPSE PRESSURE OF PALMITIC ACID TO NEARLY 70 MILLINENTONS  
CC PER METER.  
CC -1- SUBUNIT: Homodimer; disulfide-linked.  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%  
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,  
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL  
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).  
CC -1- SIMILARITY: Contains 1 saposin A-type domain.  
CC -1- SIMILARITY: Contains 3 saposin B-type domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X14778; CAA32865.1; -  
DR PIR; S02766; S02766.  
DR HSSP; P07988; IDFW.  
DR InterPro; IPR003119; SApA.  
DR InterPro; IPR000004; SApB.  
DR InterPro; IPR003258; Surfactant\_B.  
DR Pfam; PF02199; SApA; 1.  
DR Pfam; PF01864; SApB\_1; 1.  
DR Pfam; PF01489; SApB\_2; 3.  
DR Pfam; PF001732; SApB\_sub; 1.  
DR ProDom; PD008002; Surfactant\_B; 1.  
DR SMART; SM00162; SApA; 1.  
DR SMART; SM00118; SApB; 3.  
DR Surface film; Gaseous exchange; Glycoprotein; Repeat.  
KM

FT PROPEP 1 190  
FT CHAIN 191 269  
FT PROPEP 270 376  
FT DOMAIN 26 59  
FT DISULFID 198 267  
FT DISULFID 201 261  
FT DISULFID 225 236  
FT DISULFID 238 238  
FT CARBOHYD 306 306  
SQ SEQUENCE 376 AA; 41590 MM; F329DC62E73FB4C CRC64;  
Query Match 83.1%; Score 343; DB 1; Length 376;  
Best Local Similarity 80.5%; Pred. No. 1.1e-32;  
Matches 62; Conservative 8; Mismatches 7; Indels 0; Gaps 0;  
OY 2 PIPPLPYCMLCRALIKRIQAMIPKCALVAVAAQVGVVPLVAGGICQCLAEERYVILDTL 61  
DB 192 PIPPLPYCMLCRALIKRIQAMIPKCALVAVAAQVGVVPLVAGGICQCLAEERYVILDTL 251  
OY 62 LGRLPOLVCRVLRCS 78  
DB 252 LGRLPOLVCRVLRCS 268  
RESULT 5  
PSPB\_MOUSE STANDARD; PRT; 377 AA.  
AC P50405;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Pulmonary surfactant-associated protein B precursor (SP-B) (Pulmonary  
DE surfactant-associated proteolipid SPL (Phe)).  
GN SFTPB OR SFTP3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=DEA/20; TISSUE=Liver;  
RX MEDLINE=95208782; PubMed=7900819;  
RA Bruno M.A., Bohnsack R.J., Carter J.E., Foss K.A., Whitteatt J.A.;  
RT "Structure and function of the mouse surfactant protein B gene.";  
RL Am. J. Physiol. 268:L381-L389(1995).  
CC -1- FUNCTION: PULMONARY SURFACTANT-ASSOCIATED PROTEINS PROMOTE  
CC ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-  
CC LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES. SP-B INCREASES  
CC THE COLLAPSE PRESSURE OF PALMITIC ACID TO NEARLY 70 MILLINENTONS  
CC PER METER.  
CC -1- SUBUNIT: Homodimer; disulfide-linked.  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%  
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,  
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL  
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).  
CC -1- SIMILARITY: Contains 1 saposin A-type domain.  
CC -1- SIMILARITY: Contains 3 saposin B-type domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; S78114; AAB34846.2; -  
DR HSSP; P07988; IDFW.  
DR MGD; MGI:109516; Sftpb.  
DR InterPro; IPR003119; SApA.  
DR InterPro; IPR000004; SApB.





QY 55 VILDTLGLRMPLPOLYCRVLNCS 78  
DB 367 RSITSLIDPEASPELVCSMLHCS 390

RESULT 9  
SAP HUMAN STANDARD; PRT; 524 AA.  
ID P07602; P07692; P15793; P78538; P78541; P78546; P78547; P78558;  
AC P07602; P07692; P15793; P78538; P78541; P78546; P78547; P78558;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Proactivator polypeptide precursor [Contains: Saposin A (Protein A);  
DE Saposin B (Sphingolipid activator protein 1) (SAP-1) (Cerebroside  
DE sulfate activator) (CSAc) (Diaperin) (Sulfatide/GM1 activator);  
DE Saposin C (Co-beta-glucosidase) (AI activator) (Glucosylceramidase  
DE activator) (Sphingolipid activator protein 2) (SAP-2); Saposin D  
DE (Protein C) (Component C)].  
GN PSAP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=90129043; PubMed=2515150;  
RA Rorman E.G., Grabowski G.A.;  
RT "Molecular cloning of a human co-beta-glucosidase cDNA: evidence that  
RT four sphingolipid hydrolase activator proteins are encoded by single  
RT genes in humans and rats.";  
RL Genomics 5:486-492(1989).  
RN (2)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=8925151; PubMed=2498298;  
RA Nakano T., Sanhoff K., Stuenkel J., Christomanou H., Suzuki K.;  
RT "Structure of full-length cDNA coding for sulfatide activator, a  
RT Co-beta-glucosidase and two other homologous proteins: two alternate  
RT forms of the sulfatide activator.";  
RL J. Biochem. 105:152-154(1989).  
RN (3)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Eye, and Skin;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klauener R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udell T.B., Tohyuki S., Carninci P., Prange C.,  
RA Raha S.S., Lonnellano N.A., Peters G.J., Abramson R.D., Mulhany S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richard S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiy S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Maria M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16699-16903(2002).  
RN (4)  
RP SEQUENCE OF 59-125 AND 304-513 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=91192146; PubMed=2013321;  
RA Holtschmidt H., Sandhoff K., Fuert W., Kwon H.Y., Schnabel D.,  
RA Suzuki K.;  
RT "The organization of the gene for the human cerebroside sulfate  
RT activator protein.";

RL FEBS Lett. 280:267-270(1991).  
RN (5)  
RP SEQUENCE OF 164-524 FROM N.A.  
RX MEDLINE=88068647; PubMed=2825202;  
RA Dewji N.N., Wenger D.A., O'Brien J.S.;  
RT "Nucleotide sequence of cloned cDNA for human sphingolipid activator  
RT protein 1 precursor.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:8652-8656(1987).  
RN (6)  
RP PARTIAL SEQUENCE OF 60-142.  
RX MEDLINE=89240739; PubMed=2717620;  
RA Morimoto S., Martin B.M., Yamamoto Y., Kretz K.A., O'Brien J.S.,  
RA Kishimoto Y.;  
RT "Saposin A: second cerebroside activator protein.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:3389-3393(1989).  
RN (7)  
RP SEQUENCE OF 195-263 FROM N.A.  
RX MEDLINE=86130593; PubMed=2868718;  
RA Dewji N.N., Wenger D.A., Fujibayashi S., Donoviel M., Bach F.,  
RA Hill F., O'Brien J.S.;  
RT "Molecular cloning of the sphingolipid activator protein-1 (SAP-1),  
RT the sulfatide sulfate activator.";  
RL Biochem. Biophys. Res. Commun. 134:989-994(1986).  
RN (8)  
RP SEQUENCE OF 195-274.  
RC TISSUE=Kidney;  
RX MEDLINE=91006165; PubMed=2209618;  
RA Fuert W., Schuber J., Machleidt W., Meyer H.E., Sandhoff K.;  
RT "The complete amino acid sequences of human ganglioside GM2 activator  
RT protein and cerebroside sulfate activator protein.";  
RL Eur. J. Biochem. 192:709-714(1990).  
RN (9)  
RP SEQUENCE OF 195-274.  
RX MEDLINE=89207118; PubMed=3242555;  
RA Kleinschmidt T., Christomanou H., Braunitzer G.;  
RT "Complete amino acid sequence of the naturally occurring A2 activator  
RT protein for enzymic sphingomyelin degradation: identity to the  
RT sulfatide activator protein (SAP-1)."  
RL Biol. Chem. Hoppe-Seyler 369:1361-1365(1988).  
RN (10)  
RP SEQUENCE OF 311-390.  
RX MEDLINE=88163077; PubMed=3442600;  
RA Kleinschmidt T., Christomanou H., Braunitzer G.;  
RT "Complete amino acid sequence and carbohydrate content of the  
RT naturally occurring glucosylceramide activator protein (AI activator)  
RT absent from a new human Gaucher disease variant.";  
RL Biol. Chem. Hoppe-Seyler 368:1571-1578(1987).  
RN (11)  
RP SEQUENCE OF 407-484.  
RX MEDLINE=89000190; PubMed=3048308;  
RA Fuert W., Machleidt W., Sandhoff K.;  
RT "The precursor of sulfatide activator protein is processed to three  
RT different proteins.";  
RL Biol. Chem. Hoppe-Seyler 369:317-328(1988).  
RN (12)  
RP PARTIAL SEQUENCE OF 405-484.  
RX MEDLINE=89025876; PubMed=2845979;  
RA Morimoto S., Martin B.M., Kishimoto Y., O'Brien J.S.;  
RT "Saposin D: a sphingomyelinase activator.";  
RL Biochem. Biophys. Res. Commun. 156:403-410(1988).  
RN (13)  
RP SEQUENCE OF 17-26.  
RC TISSUE=Milk;  
RX MEDLINE=92068206; PubMed=1958198;  
RA Kondoh K., Hino T., Sano A., Kakimoto Y.;  
RT "Isolation and characterization of prosaposin from human milk.";  
RL Biochem. Biophys. Res. Commun. 181:286-292(1991).  
RN (14)  
RP PARTIAL SEQUENCE (SAPOSIN B), AND STRUCTURE OF CARBOHYDRATES.  
RC TISSUE=Urine;  
RX MEDLINE=20032116; PubMed=10562467;  
RA Fluharty A.L., Lombardo C., Louis A., Stevens R.L., Whitelegge J.,  
RA Waring A.J., To T., Fluharty C.B., Fauli K.F.;



RT "Preparation of the cerebroside sulfate activator (CSact or saposin B)  
from human urine.";  
RL Mol. Genet. Metab. 68:391-403(1999).  
RN [15]  
RP STRUCTURE OF CARBOHYDRATE ON ASN-215.  
RX MEDLINE=21110404; PubMed=1180632;  
RA Falli K.F., Johnson J., Kim M.J., To T., Whitelegge J.P.,  
Stevens R.L., Fluharty C.B., Fluharty A.L.;  
RT "Structure of the asparagine-linked sugar chains of porcine kidney and  
human urine cerebroside sulfate activator protein.";  
RL J. Mass Spectrom. 35:1416-1424(2000).  
RN [16]  
RP SAPOSIN D DISULFIDE BONDS.  
RX MEDLINE=99337689; PubMed=10406959;  
RA Tatti M., Salvio R., Ciaffoni F., Pucci P., Andolfo A.,  
Amoresano A., Vaccaro A.M.;  
RT "Structural and membrane-binding properties of saposin D.";  
RL Eur. J. Biochem. 263:486-494(1999).  
RN [17]  
RP MASS SPECTROMETRY.  
RX TISSUE=Urine;  
MEDLINE=99441404; PubMed=10510427;  
RA Falli K.F., Whitelegge J.P., Higginson J., To T., Johnson J.,  
Krutchinsky A.N., Standing K.G., Waring A.J., Stevens R.L.,  
Fluharty C.B., Fluharty A.L.;  
RT "Cerebroside sulfate activator protein (Saposin B): chromatographic  
and electrospray mass spectrometric properties.";  
RL J. Mass Spectrom. 34:1040-1054(1999).  
RN [18]  
RP VARIANT MLD ILE-217.  
RX MEDLINE=90147748; PubMed=2302219;  
RA Rafi M.A., Zhang X.-L., Degala G., Wenger D.A.;  
RT "Detection of a point mutation in sphingolipid activator protein-1  
RNA in patients with a variant form of metachromatic  
leukodystrophy.";  
RL Biochem. Biophys. Res. Commun. 166:1017-1023(1990).  
RN [19]  
RP SEQUENCE FROM N.A., AND VARIANT MLD ILE-217.  
RX MEDLINE=90207231; PubMed=2320574;  
RA Kretz K.A., Carson G.S., Morimoto S., Kishimoto Y., Fluharty A.L.,  
O'Brien J.S.;  
RT "Characterization of a mutation in a family with saposin B  
deficiency: a glycosylation site defect.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:2541-2544(1990).  
RN [20]  
RP VARIANT MLD SER-241, SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RX MEDLINE=91210267; PubMed=2019586;  
RA Holtschmidt H., Sandhoff K., Kwon H.Y., Harzer K., Nakano T.,  
Suzuki K.;  
RT "Sulfate activator protein. Alternative splicing that generates  
three mRNAs and a newly found mutation responsible for a clinical  
disease.";  
RL J. Biol. Chem. 266:7556-7560(1991).  
RN [21]  
RP VARIANT GAUCHER PHE-388.  
RX MEDLINE=91285107; PubMed=2060627;  
RA Schnabel D., Schroeder M., Sandhoff K.;  
RT "Mutation in the sphingolipid activator protein 2 in a patient with a  
variant of Gaucher disease.";  
RL FEBS Lett. 284:57-59(1991).  
RN [22]

Query Match 22.0%; Score 91; DB 1; Length 524;  
Best Local Similarity 27.6%; Pred. No. 0.0038; Indels 4; Gaps 1;  
Matches 21; Conservative 15; Mismatches 36;

QY 7 YCMCLRALIKRIQAMIPKG---ALRVAVQVCRVPLVAGGICCOCLARYSVLLDTLL 62  
DB 314 YCEVCEFLVKEVTKLIDNNKTEKILDAEDKCKSKLPKSLSECCQVVDYSSILSL 373  
QY 63 GRMLPOLYGRVLRLCS 78  
DB 374 EVSPSLVCSMLHCS 389

RESULT 10  
SAP\_RAT ID SAP\_RAT STANDARD; PRT; 554 AA  
AC P10960; Q62841; Q64190;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Sulfated glycoprotein 1 precursor (SGP-1) (Prosaposin).  
GN PSAP OR SGP1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxId=10116;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Sertoli cells;  
RX MEDLINE=89000647; PubMed=3048385;  
RA Collard M.W., Sylvester S.R., Tsuruta J.K., Griswold M.D.;  
RT "Biosynthesis and molecular cloning of sulfated glycoprotein 1  
secreted by rat Sertoli cells: sequence similarity with the  
70-kilodalton precursor to sulfatide/GM1 activator.";  
RL Biochemistry 27:4557-4564(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=96128541; PubMed=8573994;  
RA Morales C.R., El-Alfy M., Zhao Q., Igdoura S.A.;  
RT "Molecular role of sulfated glycoprotein-1 (SGP-1/prosaposin) in  
Sertoli cells.";  
RL Histol. Histopathol. 10:1023-1034(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=96175245; PubMed=8601692;  
RA Morales C.R., El-Alfy M., Zhao Q., Igdoura S.A.;  
RT "Expression and tissue distribution of rat sulfated glycoprotein-1  
(prosaposin).";  
RL J. Biochem. Cytochem. 44:327-337(1996).  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- SIMILARITY: Contains 2 saposin A-type domains.  
CC -1- SIMILARITY: Contains 4 saposin B-type domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation-  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M19936; AAA42136.1; -;  
DR EMBL; S81353; AAB36042.2; -;  
DR EMBL; S81373; AAB36233.2; -;  
DR PIR; A28716; A28716.  
DR InterPro; IPR003119; SAPA.  
DR InterPro; IPR000004; SAPB.  
DR Pfam; PF02199; SAPA; 2.  
DR Pfam; PF03489; SAPB; 1; 3.  
DR Pfam; PF03489; SAPB; 2; 4.  
DR ProDom; PD001732; SAPB sub; 3.  
DR SMART; SM00162; SAPA; 2.  
DR SMART; SM00118; SAPB; 4.  
KW Sulfation; Signal; Glycoprotein; Repeat.  
FT SIGNAL 1  
FT CHAIN 1  
FT DOMAIN 17 554 SULFATED GLYCOPROTEIN 1.  
FT DOMAIN 21 54 SAPOSIN-LIKE TYPE A 1.  
FT DOMAIN 61 138 SAPOSIN-LIKE TYPE B 1.  
FT DOMAIN 193 274 SAPOSIN-LIKE TYPE B 2.  
FT DOMAIN 310 391 SAPOSIN-LIKE TYPE B 3.  
FT DOMAIN 435 516 SAPOSIN-LIKE TYPE B 4.  
FT DOMAIN 521 554 SAPOSIN-LIKE TYPE A 2.

FT DISULFID 63 138 BY SIMILARITY.  
FT DISULFID 66 132 BY SIMILARITY.  
FT DISULFID 94 106 BY SIMILARITY.  
FT DISULFID 197 270 BY SIMILARITY.  
FT DISULFID 200 264 BY SIMILARITY.  
FT DISULFID 229 240 BY SIMILARITY.  
FT DISULFID 314 387 BY SIMILARITY.  
FT DISULFID 317 381 BY SIMILARITY.  
FT DISULFID 345 356 BY SIMILARITY.  
FT DISULFID 439 512 BY SIMILARITY.  
FT DISULFID 442 506 BY SIMILARITY.  
FT DISULFID 470 481 BY SIMILARITY.  
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 115 115 P -> L (IN REF. 2 AND 3).  
FT CONFLICT 299 299 I -> E (IN REF. 2).  
FT CONFLICT 462 462 I -> V (IN REF. 3).  
FT CONFLICT 527 527 W -> R (IN REF. 3).  
FT CONFLICT 536 536 S -> M (IN REF. 3).  
SQ SEQUENCE 554 AA; 61123 MM; DE3F3A3A0520C6B CRC64;

Query Match 21.1%; Score 87; DB 1; Length 554;  
Best Local Similarity 23.7%; Pred. No. 0.012;  
Matches 18; Conservative 21; Mismatches 33; Indels 4; Gaps 1;

OY 7 YCWLRLRIKRIQMIPIKGA---LRVAAGCVPLVAGGICQCLAEKRYSVLLDTLL 62  
DB 313 FCOVCQVWKKLSELLNNNTTELLINGLSKACSLPAPASTKCEVLVTGPELDDVLM 372  
OY 63 GRMLPOLVCRILVLRCS 78  
DB 373 HEVNPFLCGVISLCS 388

RESULT 11  
SAP\_CAVPO STANDARD; PRT; 81 AA.  
ID SAP\_CAVPO  
AC P20097;  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Saposin C (CO-beta-glucosidase) (Glucosylceramidase activator)  
OS (Sphingolipid activator protein 2) (SAP-2).  
OC Cavia porcellus (Guinea pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.  
OX NCBI\_TaxID=10141;  
RN [1]  
RP SEQUENCE.  
RC TISSUE=Liver;  
RX MEDLINE=9066787; PubMed=1198642;  
RA Sano A., Radin N.S., Johnson L.L., Tarr G.E.;  
RT "The activator protein for glucosylceramide beta-glucosidase from  
RT guinea pig liver. Improved isolation method and complete amino acid  
RT sequence.";  
RL J. Biol. Chem. 263:19597-19601(1988).  
CC -1- FUNCTION: SAPOSIN A AND SAPOSIN C STIMULATE THE HYDROLYSIS OF  
CC GLUCOSYLCERAMIDE BY BETA-GLUCOSYLCERAMIDASE (EC 3.2.1.45) AND  
CC GALACTOSYLCERAMIDE BY BETA-GALACTOSYLCERAMIDASE (EC 3.2.1.46).  
CC SAPOSIN-C APPARENTLY ACTS BY COMBINING WITH THE ENZYME AND ACIDIC  
CC LIPID TO FORM AN ACTIVATED COMPLEX, RATHER THAN BY SOLUBILIZING  
CC THE SUBSTRATE.  
CC -1- SIMILARITY: Contains 1 saposin B-type domain.  
DR HSSP; P42210; IQDM.  
DR InterPro; IPR000004; SAPB.  
DR Pfam; PF05184; SAPB\_1; 1.  
DR Pfam; PF03489; SAPB\_2; 1.  
DR ProDom; PD001732; SAPB\_sub; 1.  
DR SMART; SM00118; SAPB\_1.  
KW Glycoprotein; sphingolipid metabolism.

FT DOMAIN 1 81 SAPOSIN-LIKE TYPE B.  
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .).  
FT DISULFID 5 78 BY SIMILARITY.  
FT DISULFID 8 72 BY SIMILARITY.  
FT DISULFID 36 47 BY SIMILARITY.  
SQ SEQUENCE 81 AA; 8852 MM; E564CE1F0A292596 CRC64;

Query Match 20.1%; Score 83; DB 1; Length 81;  
Best Local Similarity 26.5%; Pred. No. 0.0049;  
Matches 16; Conservative 14; Mismatches 32; Indels 4; Gaps 1;

OY 8 CWLRLRIKRIQMIPIKGA---AAGCVPLVAGGICQCLAEKRYSVLLDTLLG 63  
DB 5 CKACEYVVKKVMELIDNNRTREKIHLDSCVCLPESVEQGVVDITGDSIVALLQ 64  
OY 64 RMLPOLVC 71  
DB 65 EMBSELVC 72

RESULT 12  
SAP\_CHICK STANDARD; PRT; 518 AA.  
ID SAP\_CHICK  
AC O13035;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 26-FEB-2003 (Rel. 41, Last annotation update)  
DE Proactivator polypeptide precursor [Contains: Saposin A, Saposin B,  
DE Saposin C, Saposin D].  
OS PSAP.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A. AND SEQUENCE OF 194-203.  
RC TISSUE=Brain, and liver;  
RX MEDLINE=98129745; PubMed=9461526;  
RA Azuma N., Seo H.-C., Lie O., Fu Q., Gould R.M., Hiraiwa M., Burt D.W.,  
RA Paton I.R., Morrice D.R., O'Brien J.S., Kishimoto Y.;  
RT "Cloning, expression and map assignment of chicken prosaposin.";  
RL Biochem. J. 330:321-327(1998).  
RN [2]  
RP SEQUENCE FROM N.A.

RA Altman N., Horowitz M.;  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: THE LYSOSOMAL DEGRADATION OF SPHINGOLIPIDS TAKES PLACE  
CC BY THE SEQUENTIAL ACTION OF SPECIFIC HYDROLASES. SOME OF THESE  
CC ENZYMES REQUIRE SPECIFIC LOW-MOLECULAR MASS, NON-ENZYMATIC PROTEINS:  
CC THE SPHINGOLIPIDS ACTIVATOR PROTEINS (GOPROTEINS) (BY SIMILARITY).  
CC -1- FUNCTION: SAPOSIN A AND SAPOSIN C STIMULATE THE HYDROLYSIS OF  
CC GLUCOSYLCERAMIDE BY BETA-GLUCOSYLCERAMIDASE (EC 3.2.1.45) AND  
CC GALACTOSYLCERAMIDE BY BETA-GALACTOSYLCERAMIDASE (EC 3.2.1.46).  
CC SAPOSIN-C APPARENTLY ACTS BY COMBINING WITH THE ENZYME AND ACIDIC  
CC LIPID TO FORM AN ACTIVATED COMPLEX, RATHER THAN BY SOLUBILIZING  
CC THE SUBSTRATE (BY SIMILARITY).  
CC -1- FUNCTION: SAPOSIN B STIMULATES THE HYDROLYSIS OF GALACTO-  
CC CEROSIDE SULFATE BY ARYL SULFATASE A (EC 3.1.6.8), GM1  
CC GANGLIOSIDES BY BETA-GALACTOSIDASE (EC 3.2.1.23) AND  
CC GLAUCOTRIACOSYLCERAMIDE BY ALPHA-GALACTOSIDASE A (EC 3.2.1.22).  
CC SAPOSIN-B FORMS A SOLUBILIZING COMPLEX WITH THE SUBSTRATES OF THE  
CC SPHINGOLIPID HYDROLASES (BY SIMILARITY).  
CC -1- FUNCTION: SAPOSIN D IS A SPECIFIC SPHINGOMYELIN PHOSPHODIESTERASE  
CC ACTIVATOR (EC 3.1.4.12) (BY SIMILARITY).  
CC -1- SUBUNIT: SAPOSIN B IS A HOMODIMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Lysosomal (By similarity).  
CC -1- PTM: THIS PRECURSOR IS PROTEOLYTICALLY PROCESSED TO 4 SMALL  
CC PEPTIDES, WHICH ARE SIMILAR TO EACH OTHER AND ARE SPHINGOLIPID  
CC HYDROLASE ACTIVATOR PROTEINS (BY SIMILARITY).  
CC -1- SIMILARITY: Contains 2 saposin A-type domains.  
CC -1- SIMILARITY: Contains 4 saposin B-type domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
CC EMBL; AB003471; BAA19914.1; -  
DR EMBL; AF108656; AAF05899.1; -  
DR InterPro; IPR003119; SAPA.  
DR InterPro; IPR000004; SAPB.  
DR Pfam; PF02199; SAPA; 2.  
DR Pfam; PF05184; SAPB; 1; 4.  
DR Pfam; PF03489; SAPB; 2; 4.  
DR Pfam; PF001132; SAPB; sub; 3.  
DR Pfam; PF001132; SAPB; sub; 3.  
KW Signal; Glycoprotein; Lysosome; Sphingolipid metabolism; Repeat;  
KW GM2-gangliosidosis.  
FT SIGNAL 1 17 POTENTIAL.  
FT PROPEP 18 60  
FT CHAIN 61 143 SAPOSIN A.  
FT PROPEP 145 193  
FT CHAIN 194 276 SAPOSIN B.  
FT PROPEP 278 305  
FT CHAIN 307 387 SAPOSIN C.  
FT PROPEP 389 398  
FT CHAIN 399 480  
FT PROPEP 482 518 SAPOSIN D.  
FT DOMAIN 22 55 SAPOSIN-LIKE TYPE A 1.  
FT DOMAIN 60 143 SAPOSIN-LIKE TYPE B 1.  
FT DOMAIN 193 277 SAPOSIN-LIKE TYPE B 2.  
FT DOMAIN 307 388 SAPOSIN-LIKE TYPE B 3.  
FT DOMAIN 399 480 SAPOSIN-LIKE TYPE B 4.  
FT DOMAIN 485 518 SAPOSIN-LIKE TYPE A 2.  
FT DOMAIN 64 139 BY SIMILARITY.  
FT DISULFID 67 133 BY SIMILARITY.  
FT DISULFID 95 107 BY SIMILARITY.  
FT DISULFID 197 273 BY SIMILARITY.  
FT DISULFID 200 267 BY SIMILARITY.  
FT DISULFID 229 240 BY SIMILARITY.  
FT DISULFID 311 384 BY SIMILARITY.  
FT DISULFID 314 378 BY SIMILARITY.  
FT DISULFID 342 353 BY SIMILARITY.  
FT DISULFID 403 476 BY SIMILARITY.  
FT DISULFID 406 470 BY SIMILARITY.  
FT DISULFID 434 445 BY SIMILARITY.  
FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 420 420 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 94 94 R -> T (IN REF. 2).  
FT CONFLICT 486 486 E -> D (IN REF. 2).  
SQ SEQUENCE 518 AA; 57601 MW; B803000E891C3963 CRC64;

Query Match 19.5%; Score 80.5; DB 1; Length 518;  
Best Local Similarity 24.3%; Pred. No. 0.064;  
Matches 18; Conservative 15; Mismatches 36; Indels 5; Gaps 2;

QY 8 CMLCALIKRIQAMIPKGLRVAVAQVCR-----VCRVPLVAGGICQCLAERYSYLLDTLLG 63  
DB 311 CEICETWAEVTVGLLESNKTREEIVHEWVCVLLPASVYKQCKDFIEYGGALIDMLE 370

QY 64 RMLPOLVCRVLVRC 77  
DB 371 ATNPEAVC-VMLKC 383

RESULT 13  
FX24 HUMAN STANDARD; PRT; 318 AA.  
ID FX24 HUMAN  
AC 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE F-box only protein 24.  
GN FBX024 OR FBX24.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gloeckner G., Rosenchal A., Scherer S., Weber J., Schattevoy R.,  
RA Tsai L.-C.;  
RT "Large scale analysis of two regions in human chromosome 7q22:  
RT annotation of 650 kb of genomic sequence around the PCOLCE and CUTL1  
RT loci reveals 17 genes";  
RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20003060; PubMed=10531035;  
RA Cenciarelli C., Chiau D.S., Guardavaccaro D., Parks W., Vidal M.,  
RA Pagano M.;  
RL "Identification of a family of human F-box proteins";  
RL Curr. Biol. 9:1177-1179(1999).  
CC -1- FUNCTION: Probably recognizes and binds to some phosphorylated  
CC proteins and promotes their ubiquitination and degradation.  
CC -1- SUBUNIT: PART OF A SCF (SKP1-CULLIN-F-BOX) PROTEIN LIGASE COMPLEX  
CC (BY SIMILARITY).  
CC -1- SIMILARITY: Contains 1 F-box domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
CC EMBL; AF053356; AAC78801.1; -  
DR EMBL; AF174604; AAF04525.1; -  
DR Genbank; HGNC:13595; FBX024.  
DR GO; GO:0000153; C:ubiquitin ligase complex; NAS.  
DR GO; GO:0004842; F:ubiquitin-protein ligase activity; NAS.  
DR GO; GO:0016567; P:protein ubiquitination; NAS.  
DR InterPro; IPR001810; F-box.  
DR Pfam; PF00646; F-box; 1.  
DR SMART; SM00256; FBX01; 1.  
DR PROSITE; PS00181; FBX01; 1.  
KW Ubl conjugation.  
KW DOMAIN 36 82 F-BOX.  
SQ SEQUENCE 318 AA; 36353 MW; 025D9AF7342646F2 CRC64;

Query Match 16.0%; Score 66; DB 1; Length 318;  
Best Local Similarity 29.1%; Pred. No. 1.9;  
Matches 25; Conservative 10; Mismatches 33; Indels 18; Gaps 2;

QY 10 LCRALIKRIQAMIPKGLRVAVAQVCR-----VPLVAGGICQCLAERYSYLLD----- 59  
DB 82 ICRSPRLDQDPRKGLVFGAFGRRRLCKSVAPLHAGRRFLPTDHFILDYVCTL 141

QY 60 -----TLGRMLPOLVCRVLVRC 77  
DB 142 FFLKNAVLSTLGGOMQWKRACRYVVLIC 167

RESULT 14  
ACV5 CEPAC STANDARD; PRT; 3712 AA.  
ID ACV5 CEPAC  
AC P25464;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE N-(5-amino-5-carboxypentanoyl)-L-cysteiny]-D-valine synthase  
DE (EC 6.3.2.26) (Delta-(L-alpha-aminoadipyl)-L-cysteiny]-D-valine  
DE synthetase) (ACV synthetase) (ACVS).

GN PCBAB.  
 OS Cephalosporium acremonium (Acremonium chrysogenum).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreomycetidae; Hypocreales; Hypocreaceae; mitosporic Hypocreaceae;  
 OC Acremonium.  
 RN NCBI\_TaxID=5044;  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91177827; PubMed=1706706;  
 RA Gutierrez S., Diez B., Montenegro B., Martin J.F.;  
 RT "Characterization of the Cephalosporium acremonium pcbab gene  
 RT encoding alpha-aminoadipyl-L-cysteinyl-valine synthetase, a large  
 RT multidomain peptide synthetase: linkage to the pcbC gene as a cluster  
 RT of early cephalosporin biosynthetic genes and evidence of multiple  
 RT functional domains.";  
 RL J. Bacteriol. 173:2354-2365(1991).  
 RN [2]  
 RN PARTIAL SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RP STRAIN=ATCC 11550;  
 RX MEDLINE=91168300; PubMed=2076552;  
 RA Hoskins J.A., O'Callaghan N., Queener S.W., Cantwell C.A., Wood J.S.,  
 RA Chen V.J., Skarud P.L.;  
 RT "Gene disruption of the pcbab gene encoding ACV synthetase in  
 RT Cephalosporium acremonium.";  
 RL Curr. Genet. 18:523-530(1990).  
 CC -1- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE  
 CC ACV ARE ACTIVATED AS AMINOACYL-ADENYLATES WITH PEPTIDE BONDS  
 CC FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER  
 CC INTERMEDIATES.  
 CC -1- CATALYTIC ACTIVITY: L-2-aminohexanediolate + L-cysteine + L-valine  
 CC + 3 ATP = N-[L-5-amino-5-carboxypentanoyl]-L-cysteinyl-D-valine +  
 CC 3 AMP + 3 diphosphate.  
 CC -1- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES  
 CC (POTENTIAL).  
 CC -1- PATHWAY: Biosynthesis of penicillin and cephalosporin, first step.  
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME  
 CC FAMILY.  
 CC -1- SIMILARITY: Contains 3 acyl carrier domains.  
 CC PIR: A38531; YGCEVC.  
 DR HSSP: P14687; 1AMU.  
 DR InterPro: IPR000873; AMP-bind.  
 DR InterPro: IPR001242; Condensatn.  
 DR InterPro: IPR006163; PP\_bind.  
 DR InterPro: IPR006162; Pantene\_attach.  
 DR InterPro: IPR000379; Ser\_estrs\_site.  
 DR InterPro: IPR001031; Thioesterase.  
 DR Pfam: PF00501; AMP-binding; 3.  
 DR Pfam: PF00668; Condensation; 3.  
 DR Pfam: PF00550; PP-binding; 3.  
 DR Pfam: PF00975; Thioesterase; 1.  
 DR PRINTS: PR00154; AMPBINDING.  
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 2.  
 DR PROSITE: PS00455; AMP\_BINDING; 3.  
 DR PROSITE: PS00075; ACP\_DOMAIN; 3.  
 KW Ligase; Antibiotic biosynthesis; Multifunctional enzyme;  
 KW Repeat; Phosphopantetheine.  
 FT REPEAT 234 1062 DOMAIN 1 (ADIPATE-ACTIVATING).  
 FT REPEAT 1335 2162 DOMAIN 2 (CYSTEINE-ACTIVATING).  
 FT REPEAT 2409 3387 DOMAIN 3 (VALINE-ACTIVATING).  
 FT DOMAIN 795 864 ACYL\_CARRIER (ACP) 1.  
 FT DOMAIN 1880 1953 ACYL\_CARRIER (ACP) 2.  
 FT DOMAIN 2960 3027 ACYL\_CARRIER (ACP) 3.  
 FT BINDING 827 827 PHOSPHOPANTETHEINE (BY SIMILARITY).  
 FT BINDING 1916 1916 PHOSPHOPANTETHEINE (BY SIMILARITY).  
 FT BINDING 2990 2990 PHOSPHOPANTETHEINE (BY SIMILARITY).  
 FT ACT\_SITE 3568 3568 THIOESTERASE (BY SIMILARITY).  
 SQ SEQUENCE 3712 AA; 414767 MW; 4EE3CIEB5EBF9B7 CRC64;

Query Match 15.4%; Score 63.5; DB 1; Length 3712;  
 Best Local Similarity 37.2%; Pred. No. 46;  
 Matches 16; Conservative 6; Mismatches 20; Indels 1; Gaps 1;  
 QY 27 LRVAAGVCRVPLVAGGICQCLAEKRYSVLLDTLLGRMLPOL 69

Db 3237 LRRTVAAGSEFTLPCMGVLCQ-QEKFSTRRTTALLSKAPAL 3278  
 RESULT 15  
 ID RAFR\_PEDPE STANDARD; PRT; 277 AA.  
 AC P43465;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DE 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Rafinose operon transcriptional regulatory protein rafr.  
 GN RAFR.  
 OS Pedicoccus pentosaceus.  
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae; Pedicoccus.  
 OX NCBI\_TaxID=1255;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=PEP1.0;  
 RA Leenhouts K.K.J., Bolhuis A.A., Kok J.J., Venema G.G.;  
 RL Submitted (XXX-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: INVOLVED IN THE REGULATION OF THE RAFINOSE-OPERON.  
 CC -1- SIMILARITY: BELONGS TO THE ARAC/XYLS FAMILY OF TRANSCRIPTIONAL  
 CC REGULATORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: Z32771; CA83663.1; -  
 DR EMBL: L32093; AAA25562.1; -  
 DR PIR: S44252; S44252.  
 DR InterPro: IPR000005; HTHARAC.  
 DR Pfam: PF00165; HTH\_ARAC; 2.  
 DR SMART: SM00342; HTH\_ARAC; 1.  
 DR PROSITE: PS00041; HTH\_ARAC\_FAMILY\_1; 1.  
 DR PROSITE: PS01124; HTH\_ARAC\_FAMILY\_2; 1.  
 KW Transcription regulation; Activator; DNA-binding.  
 FT DNA\_BIND 192 211 H-T-H MOTIF (BY SIMILARITY).  
 SQ SEQUENCE 277 AA; 31532 MW; 8CE8A5E9412B200 CRC64;

Query Match 15.3%; Score 63; DB 1; Length 277;  
 Best Local Similarity 31.9%; Pred. No. 3.7;  
 Matches 23; Conservative 14; Mismatches 25; Indels 10; Gaps 3;  
 QY 4 PLPYCWLCAALIKRIQAMTPKGLRVAVAGVPLVAGGICQCLAEKRYSV-----IL 57  
 Db 85 PMRYFMIGFSAGIRIAMU--SGSL--LAQKCYLRQVNGHIYADSELYKVLHIPSLLI 140  
 QY 58 LDTLLGRMLPOL 69  
 Db 141 NDVLGSLYRL 152

Search completed: December 3, 2003, 15:50:32  
 Job time: 14.9728 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 15:44:20 / Search time 56.4286 Seconds  
(without alignments)  
361.274 Million cell updates/sec

361.274 Million cell updates/sec

Title: US-09-788-308d-2

Sequence: 1 PPIPLPYCWLGRALIKRIQA.....TLGRMLPOLVCRVLRCSSM 79

Scoring table: BL0SUM62  
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	356	86.2	228	6	Q9BD29
2	356	86.2	243	6	Q9T05
3	356	86.2	374	6	Q9T05
4	340	82.3	370	6	Q9T05
5	340	82.3	378	11	P79333
6	248.5	60.2	205	6	Q9N275
7	104	25.2	520	13	Q9N275
8	104	25.2	522	13	Q9N275
9	93.5	22.6	965	5	Q15997
10	90.5	21.9	307	5	Q9BK02
11	87.5	21.2	294	5	Q9BK02
12	81.5	19.7	97	6	Q9BK02
13	78.5	19.0	453	4	Q9BK02
14	76	18.4	81	6	Q9BK02
15	73.5	17.8	241	4	Q9BK02
16	73	17.7	469	5	Q9BK02

17	73	17.7	953	5	Q9Y125
18	71	17.2	441	5	Q9Y125
19	70	16.9	458	5	Q9Y125
20	70	16.9	484	5	Q9BK01
21	68.5	16.6	402	11	Q9BK01
22	68.5	16.6	449	11	Q9BK01
23	68.5	16.6	525	11	Q9BK01
24	65	15.7	456	5	Q9Y125
25	65	15.7	456	5	Q9Y125
26	65	15.7	456	5	Q9Y125
27	65	15.7	456	5	Q9Y125
28	64.5	15.6	446	8	Q9Y125
29	64	15.5	233	17	Q9Y125
30	64	15.5	281	17	Q9Y125
31	63.5	15.4	1407	3	Q06211
32	63	15.3	554	11	Q9BK01
33	63	15.3	779	13	Q9Y125
34	62.5	15.1	157	5	Q9Y125
35	62.5	15.1	446	8	Q9Y125
36	62.5	15.1	446	8	Q9Y125
37	62.5	15.1	446	8	Q9Y125
38	62.5	15.1	446	8	Q9Y125
39	62.5	15.1	446	8	Q9Y125
40	62.5	15.1	446	8	Q9Y125
41	62.5	15.1	446	8	Q9Y125
42	62.5	15.1	446	8	Q9Y125
43	62.5	15.1	555	2	Q9Y125
44	62.5	15.1	731	5	Q9Y125
45	62	15.0	336	10	Q9Y125

## ALIGNMENTS

Q9BD29	PRELIMINARY;	PRT;	228 AA.
Q9BD29	Q9BD29		
AC	Q9BD29		
DT	01-JUN-2001 (TEMBLrel. 17, Created)		
DT	01-JUN-2001 (TEMBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TEMBLrel. 21, Last annotation update)		
DE	Pulmonary surfactant-associated protein B precursor (Fragment).		
GN	SP-B.		
OS	Ovis aries (Sheep).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Eucleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
OC	Bovidae; Caprinae; Ovis.		
OX	NCBI_TaxID=9940;		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Lung;		
RA	Pieteschmann S.M., Pison U.;		
RT	"CDNA sequence of ovine pulmonary surfactant-associated protein B (SP-B)."		
RT	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.		
RL	EMBL: AF076635; AKL14977.1; -		
DR	HSSP: P07988; IDFW.		
DR	InterPro: IPR000004; SAPB.		
DR	InterPro: IPR003258; Surfactant_B.		
DR	Pfam: PF03489; Surfactant_B; 1.		
DR	ProDom: PD008002; Surfactant_B; 1.		
DR	SMART: SM00118; SAPB; 2.		
KW	Signal.		
FT	NON_TER		
FT	STGNL		
FT	CHAIN		
FT	SEQUENCE		
SO	228 AA; 24766 MW; C6A02A6CDFD9006 CRC64;		

Query Match 86.2%; Score 356; DB 6; Length 228;  
Best Local Similarity 84.6%; Pred. No. 4e-34;  
Matches 66; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

```
QY 1 FPIPLPYCMLCRALIKRIQAMIPKCALRVAVAVQCRVPLVAGGICQCLAEERYVILLDT 60
DB 43 FPIPLPYCMLCRALIKRIQAMIPKCALRVAVAVQCRVPLVAGGICQCLAEERYVILLDT 102
QY 61 LLGRMLPOLVCRVLVRCSS 78
DB 103 LLGRMLPOLVCRVLVRCSS 120

RESULT 2
Q9TU05 PRELIMINARY; PRT; 243 AA.
AC Q9TU05;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Pulmonary surfactant protein B (Fragment).
GN SPB5.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OC NCBI_TaxId=9940;
RN (1)
RP SEQUENCE FROM N.A.
RA Braems G.A., Yao L.-D., Inchley K., Brickenden A., Han V.K.M.,
RA Grolla A., Challis J.R.G., Possmayer F.;
RT "cDNAs for ovine surfactant proteins: application in studies on fetal
RT lung growth and maturation following prolonged hypoxemia.";
RL Submitted (DEC-1999) to the EMBL/Genbank/DBD databases.
DR EMBL; AF211857; AAF18996.1; -.
DR HSSP; P07988; 1DFW.
DR InterPro; IPR000004; SAPB.
DR InterPro; IPR003258; Surfactant_B.
DR Pfam; PF03489; Surfactant_B; 1.
DR ProDom; PD008002; Surfactant_B; 1.
DR SMART; SM00118; SAPB; 2.
DR NON TER
FT
SQ SEQUENCE 243 AA; 26472 MW; 7DC15FOABA88ED2 CRC64;

Query Match 86.2%; Score 356; DB 6; Length 243;
Best Local Similarity 84.6%; Pred. No. 4,3e-34;
Matches 66; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 FPIPLPYCMLCRALIKRIQAMIPKCALRVAVAVQCRVPLVAGGICQCLAEERYVILLDT 60
DB 58 FPIPLPYCMLCRALIKRIQAMIPKCALRVAVAVQCRVPLVAGGICQCLAEERYVILLDT 117
QY 61 LLGRMLPOLVCRVLVRCSS 78
DB 118 LLGRMLPOLVCRVLVRCSS 135

RESULT 3
Q9TU01 PRELIMINARY; PRT; 374 AA.
AC Q9TU01;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Pulmonary surfactant-associated protein B (Fragment).
GN Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OC NCBI_TaxId=9940;
RN (1)
RP SEQUENCE FROM N.A.
RA TISSUE=Lung;
RX MEDLINE=20215263; PubMed=10749754;
RA Pletschmann S.M., Pison U.;
RT "cDNA cloning of ovine pulmonary surfactant proteins A, B, and C:
RT isolation of two different sequences for Sp-B.";
```

```
RL Am. J. Physiol. 278:L765-L778 (2000).
DR EMBL; AF107544; AAF14195.1; -.
DR HSSP; P07988; 1DFW.
DR InterPro; IPR003119; SAPA.
DR InterPro; IPR000004; SAPB.
DR InterPro; IPR003258; Surfactant_B.
DR Pfam; PF02199; SAPA; 1.
DR Pfam; PF03489; Surfactant_B; 1.
DR ProDom; PD001732; SAPB; 1.
DR ProDom; PD008002; Surfactant_B; 1.
DR SMART; SM00162; SAPA; 1.
DR SMART; SM00118; SAPB; 3.
DR NON TER
FT
SQ SEQUENCE 374 AA; 41291 MW; 16A4D62804AD5A2F CRC64;

Query Match 86.2%; Score 356; DB 6; Length 374;
Best Local Similarity 84.6%; Pred. No. 6,4e-34;
Matches 66; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 FPIPLPYCMLCRALIKRIQAMIPKCALRVAVAVQCRVPLVAGGICQCLAEERYVILLDT 60
DB 189 FPIPLPYCMLCRALIKRIQAMIPKCALRVAVAVQCRVPLVAGGICQCLAEERYVILLDT 248
QY 61 LLGRMLPOLVCRVLVRCSS 78
DB 249 LLGRMLPOLVCRVLVRCSS 266

RESULT 4
P79333 PRELIMINARY; PRT; 370 AA.
ID P79333;
AC P79333;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Surfactant protein B (SP-B).
GN SP-B.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OC NCBI_TaxId=9986;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=96198312; PubMed=8928820;
RA Margana R.K., Boggaram V.;
RT "Rabbit surfactant protein B gene: structure and functional
RT characterization of the promoter.";
RL Am. J. Physiol. 270:L601-L612 (1996).
DR EMBL; U40853; AAB48076.1; -.
DR HSSP; P07988; 1DFW.
DR InterPro; IPR003119; SAPA.
DR InterPro; IPR000004; SAPB.
DR InterPro; IPR003258; Surfactant_B.
DR Pfam; PF02199; SAPA; 1.
DR Pfam; PF03489; Surfactant_B; 1.
DR ProDom; PD001732; SAPB; 1.
DR ProDom; PD008002; Surfactant_B; 1.
DR SMART; SM00162; SAPA; 1.
DR SMART; SM00118; SAPB; 3.
DR NON TER
FT
SQ SEQUENCE 370 AA; 40535 MW; B23690B95C3C5138 CRC64;

Query Match 83.3%; Score 344; DB 6; Length 370;
Best Local Similarity 82.1%; Pred. No. 1,6e-32;
Matches 64; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 FPIPLPYCMLCRALIKRIQAMIPKCALRVAVAVQCRVPLVAGGICQCLAEERYVILLDT 60
DB 185 FPIPLPYCMLCRALIKRIQAMIPKCALRVAVAVQCRVPLVAGGICQCLAEERYVILLDT 244
QY 61 LLGRMLPOLVCRVLVRCSS 78
DB 245 LLGRMLPOLVCRVLVRCSS 262
```

```

RESULT 5
ID 035489 PRELIMINARY; PRT; 378 AA.
AC 035489;
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Surfactant protein-B.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NCBI_TaxId=10141;
NCBI_Locus=10141;
RN [1]
RP SEQUENCE FROM N.A.
RA Bingle C.D., Yuan H.T., Gowan S.;
RT "Guinea pig surfactant protein-B is differentially polyadenylated.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF031190; AAB87086.1; -.
DR HSSP; P07988; 1DFW.
DR InterPro; IPR003119; SAPA.
DR InterPro; IPR000004; SAPB.
DR InterPro; IPR003258; Surfactant_B.
DR Pfam; PF03489; Surfactant_B; 1.
DR ProDom; PD001732; SapB; 1.
DR ProDom; PD008002; Surfactant_B; 1.
DR SMART; SM00162; SAPA; 1.
DR SMART; SM00118; SAPB; 3.
SQ SEQUENCE 378 AA; 41677 MW; E3DA2E237ED401DA CRC64;

Query Match 82.3%; Score 340; DB 11; Length 378;
Best Local Similarity 79.7%; Pred. No. 5e-32;
Matches 63; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY 1 FPIPLPYCWLCKRLIKRIQAMIPKALRVAVQYCRVPLVAGGICCCCLARYSVILDT 60
DB 196 FPIPLPYCWLCKRLIKRIQAMIPKALRVAVQYCRVPLVAGGICCCCLARYSVILDT 255
QY 61 LIGRMPLPOLVCRVLRLCSM 79
DB 256 LLSHLPLPOLVCGVLRLCSM 274

RESULT 6
ID 09N275 PRELIMINARY; PRT; 205 AA.
AC 09N275;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Pulmonary surfactant-associated protein B precursor (Fragment).
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NCBI_TaxId=9940;
NCBI_Locus=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=lung;
RX MEDLINE=20215263; PubMed=10749754;
RT "CDNA cloning of ovine pulmonary SP-A, SP-B, and SP-C: isolation of
two different sequences for SP-B.";
RL Am. J. Physiol. 278:L765-L778 (2000).
DR EMBL; AF076636; AAF31150.1; -.
DR HSSP; P07988; 1DFW.
DR InterPro; IPR000004; SAPB.
DR InterPro; IPR003258; Surfactant_B.
DR Pfam; PF03489; Surfactant_B; 1.
DR ProDom; PD008002; Surfactant_B; 1.

```

```

DR SMART; SM00118; SAPB; 2.
KW Signal.
FT NON TER 1
FT SIGNAL <1 42
FT CHAIN 43 98
SQ SEQUENCE 205 AA; 22476 MW; 9E3E3B2ADBBA42PB CRC64;

Query Match 60.2%; Score 248.5; DB 6; Length 205;
Best Local Similarity 64.1%; Pred. No. 1.8e-21;
Matches 50; Conservative 2; Mismatches 3; Indels 23; Gaps 1;

QY 1 FPIPLPYCWLCKRLIKRIQAMIPKALRVAVQYCRVPLVAGGICCCCLARYSVILDT 60
DB 43 FPIPLPYCWLCKRLIKRIQAMIPKALRVAVQYCRVPLVAGGICCCCLARYSVILDT 79
QY 61 LIGRMPLPOLVCRVLRLCS 78
DB 80 LIGRMPLPOLVCGVLRLCS 97

RESULT 7
ID 08UV24 PRELIMINARY; PRT; 520 AA.
AC 08UV24;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)
DE Lysoosomal cofactor/neurotrophic factor prosaposin.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
NCBI_TaxId=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Seo H.-C., Lie O., Fjose A., O'Brien J.S., Kishimoto Y.;
RT "Cloning, expression and promoter analysis of zebrafish prosaposin.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF276996; AAL54381.1; -.
DR InterPro; IPR003119; SAPA.
DR InterPro; IPR000004; SAPB.
DR Pfam; PF02199; SAPA; 2.
DR ProDom; PD001732; SapB; 3.
DR SMART; SM00162; SAPA; 3.
DR SMART; SM00118; SAPB; 4.
SQ SEQUENCE 520 AA; 57431 MW; F9E620F84BA41CB5 CRC64;

Query Match 25.2%; Score 104; DB 13; Length 520;
Best Local Similarity 29.9%; Pred. No. 0.00049;
Matches 23; Conservative 15; Mismatches 35; Indels 4; Gaps 1;

QY 6 PYCWLCKRLIKRIQAMIPKALRVAVQYCRVPLVAGGICCCCLARYSVILDT 61
DB 310 POCALCEYWKELIENNIDOTSSEAIYQAEKVCNLPSTLTAQCKDLIETYGQAILDL 369
QY 62 LIGRMPLPOLVCRVLRLCS 78
DB 370 VQEADEKTVCSFLALCS 386

RESULT 8
ID 09DG82 PRELIMINARY; PRT; 522 AA.
AC 09DG82;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)
DE Prosaposin.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

```

OC Cyprinidae; Danio.  
 OX NCBI\_TaxId=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Altman N., Horowitz M.;  
 RT "The zebrafish prosaposin cDNA."  
 RL Submitted (NOV-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF108655; AAG32919.1; -  
 DR ZFIN; ZDB-GENE-020108-1; psep.  
 DR InterPro; IPR003119; SAPA.  
 DR InterPro; IPR000004; SAPB.  
 DR Pfam; PF02199; SAPA; 2.  
 DR ProDom; PD001732; SAPB; 3.  
 DR SMART; SM00162; SAPA; 2.  
 DR SMART; SM00118; SAPB; 4.  
 SQ SEQUENCE 522 AA; 57671 MW; D3C15A305725C1CD CRC64;

Query Match 25.2%; Score 104; DB 13; Length 522;  
 Best Local Similarity 29.9%; Pred. No. 0.00049;  
 Matches 23; Conservative 15; Mismatches 35; Indels 4; Gaps 1;

OY 6 PYCMLCRALIKRIQAMT---PKGALRVAAVQVCRVPLVAGGICQCLAERYSVLLDTL 61  
 Db 312 PQCALCEYVMKEIENMIDQTSSEAEIVQAVKVCNLPSTLTQAQCKDLIETVGAIIIDL 371  
 OY 62 LGRLM.POLVCRVLNCS 78  
 Db 372 VQADPKTVCSFLACS 388

## RESULT 9

ID 015997 PRELIMINARY; PRT; 965 AA.  
 AC 015997;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Bmp109.  
 OS Bombyx mori (Silk moth).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Dictysta; Bombycoidea;  
 OC Bombycidae; Bombyx.  
 OX NCBI\_TaxId=7091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98278844; PubMed=9611271;  
 RA Tamkun J., Chang P.-K., Li H., Natori M.;  
 RT "Molecular cloning of a cDNA encoding a silkworm protein which  
 contains the conserved BH regions of Bcl-2 family proteins."  
 RL Gene 212:287-293(1998).  
 DR EMBL; AB008449; BAA23126.1; -  
 DR InterPro; IPR003119; SAPA.  
 DR InterPro; IPR000004; SAPB.  
 DR Pfam; PF02199; SAPA; 2.  
 DR ProDom; PD001732; SAPB; 6.  
 DR SMART; SM00162; SAPA; 2.  
 DR SMART; SM00118; SAPB; 7.  
 SQ SEQUENCE 965 AA; 108825 MW; FA1A7BBE37F626078 CRC64;

Query Match 22.6%; Score 93.5; DB 5; Length 965;  
 Best Local Similarity 28.6%; Pred. No. 0.015;  
 Matches 22; Conservative 19; Mismatches 29; Indels 7; Gaps 3;

OY 8 CMLGALKRIGQAMIPKGLRVAAVQV---CRVPL--VAGGICQCLAERYSVLLDTL 61  
 Db 217 COICLDMMKQADQIQSNETQDEIKVEFGSKLIPKFAEG-CMKLADERVELIETL 275  
 OY 62 LGRLM.POLVCRVLNCS 78  
 Db 276 ASEMNPQAVCSVAGLCN 292

## RESULT 10

O9BK2 PRELIMINARY; PRT; 307 AA.  
 AC O9BK2;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Naegleria fowleri A pore-forming peptide.  
 GN PRNP-A.  
 OS Naegleria fowleri.  
 OC Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria.  
 OX NCBI\_TaxId=5763;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nickel R., Benkert C., Jacobs T., Marti T., Marciano-Cabral F.,  
 RA Leippe M.;  
 RT "Pore-forming peptides of Naegleria fowleri."  
 RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF154046; AAK21658.1; -  
 DR InterPro; IPR000004; SAPB.  
 DR ProDom; PD001732; SAPB; 2.  
 DR SMART; SM00118; SAPB; 3.  
 SQ SEQUENCE 307 AA; 33133 MW; 8503E4A755BC6DDF CRC64;

Query Match 21.9%; Score 90.5; DB 5; Length 307;  
 Best Local Similarity 27.6%; Pred. No. 0.012;  
 Matches 21; Conservative 15; Mismatches 39; Indels 1; Gaps 1;

OY 3 IPLPYCMLCRALIKR-IGAMIPKGLRVAAVQVCRVPLVAGGICQCLAERYSVLLDTL 61  
 Db 213 IPCPACLMAMELVEDEISQSVQSFVEDKLNKVCCKLPTSFSGYCASLVNQYFVLVQKL 272  
 OY 62 LGRLM.POLVCRVLNCS 77  
 Db 273 LLAVSPEKICKLVNDC 288

## RESULT 11

ID 095X03 PRELIMINARY; PRT; 294 AA.  
 AC 095X03;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Naegleria fowleri A (Fragment).  
 GN NP-A.  
 OS Naegleria fowleri.  
 OC Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria.  
 OX NCBI\_TaxId=5763;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nickel R., Leippe M.;  
 RT "Pore-forming peptides of Naegleria fowleri."  
 RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF196308; AAL01157.1; -  
 DR InterPro; IPR000004; SAPB.  
 DR ProDom; PD001732; SAPB; 2.  
 DR SMART; SM00118; SAPB; 3.  
 FT NON\_TER  
 SQ SEQUENCE 294 AA; 31895 MW; 0B67550766B5B1D8 CRC64;

Query Match 21.2%; Score 87.5; DB 5; Length 294;  
 Best Local Similarity 27.6%; Pred. No. 0.026;  
 Matches 21; Conservative 14; Mismatches 40; Indels 1; Gaps 1;

OY 3 IPLPYCMLCRALIKR-IGAMIPKGLRVAAVQVCRVPLVAGGICQCLAERYSVLLDTL 61  
 Db 200 IPCPACLMAMELVEDEISQSVQSFVEDKLNKVCCKLPTSFSGYCASLVNQYFVLVQKL 259  
 OY 62 LGRLM.POLVCRVLNCS 77  
 Db 260 LLAVSPEKICKLVNDC 275



```
RESULT 12
Q8H2Q3 PRELIMINARY; PRT; 97 AA.
ID Q8H2Q3
AC Q8H2Q3
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Anticribial peptide NK-lysin (Fragment).
GN NKL.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RA Davis B.G., Zhang G., Rush B.R., Ross C., Blecha F.;
RT "Partial cDNA sequence of equine NK-lysin.";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF538056; ANNI0122.1; -.
FT NON_TER 1
FT SEQUENCE 97 AA; 10813 MW; 997CA4F077B3C22F8 CRC64;
SQ
Query Match 19.7%; Score 81.5; DB 6; Length 97;
Best Local Similarity 25.7%; Pred. No. 0.046;
Matches 19; Conservative 19; Mismatches 31; Indels 5; Gaps 2;
QY 8 CMLCRALIKRIQMI-----PKGALRVAVAGVCRVPLVAGICOCCLAERYSVILDTLLG 63
DB 18 CMCSCRILOKLEPDLVGEQPEATINENASVCRNLGLRACKIKMTICRLISRDILAG 77
QY 64 RMLPOLVCRVLVRC 77
DB 78 KK-KPEVCVDIKLC 90
RESULT 13
Q9BXJ7 PRELIMINARY; PRT; 453 AA.
ID Q9BXJ7
AC Q9BXJ7
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Aminoless.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kalantry S., Manning S., Haub O., Tomihara-Newberger C., Lee H.-G.,
RA Fangman J., Diethe C., Manova K., Lacy E.;
RT "The Aminoless gene, essential for mouse gastrulation, encodes a
RT visceral endoderm-specific protein with an extracellular cysteine-rich
RT domain.";
RL Nat. Genet. 27:0-0(2001).
DR EMBL; AF328788; AAK2853.1; -.
DR Gene; HGNC:14604; AMN.
SQ SEQUENCE 453 AA; 47814 MW; A1P2ABCE4649BFD CRC64;
Query Match 19.0%; Score 78.5; DB 4; Length 453;
Best Local Similarity 34.8%; Pred. No. 0.44;
Matches 23; Conservative 9; Mismatches 25; Indels 9; Gaps 3;
QY 9 WLCRALIKRIQMIIPKALRVAV-----AQVCRVPLVAGICOCCLAERYSVILDTLL 62
DB 221 WICALLIQLGRRCPCACHFALRPGCCDCLGAVVLLTHGPAFL-ERYRARIIDTFL 279
QY 63 GRMLPQ 68
DB 280 G--LPQ 283
RESULT 14
```

```
Q9MYT6 PRELIMINARY; PRT; 81 AA.
ID Q9MYT6
AC Q9MYT6
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Surfactant protein B (Fragment).
GN SP-B.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Paananen R., Glumoff V., Sornunen R., Voorthout W., Hallan M.;
RT "Surfactant protein B localization in Eustachian tube.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ290968; CAB96174.1; -.
DR HSSP; P07988; 1DFW.
FT NON_TER 1
FT SEQUENCE 81 AA; 8951 MW; 030E7E3C466A372 CRC64;
SQ
Query Match 18.4%; Score 76; DB 6; Length 81;
Best Local Similarity 91.7%; Pred. No. 0.17;
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 FPIPLPYCWLCR 12
DB 70 FPIPLPYCWLCR 81
RESULT 15
Q8NT74 PRELIMINARY; PRT; 241 AA.
ID Q8NT74
AC Q8NT74
DT 01-OCT-2002 (TEMBLrel. 22, Created)
DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Hypochemical protein FLJ40379.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Mueashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto U., Isono Y.,
RA Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamori K., Sekine M., Kikuchi H., Kanda K., Wagaitsuna M.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahara K., Masuo Y., Negai K., Isegai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK097698; BNC05143.1; -.
DR InterPro; IPR003119; Sapa.
DR InterPro; IPR000004; Sapa.
DR Pfam; PF02199; SAPA; 1.
DR ProDom; PD001732; SApB; 2.
DR SMART; SM00162; SApB; 1.
DR SMART; SM00118; SApB; 2.
KW Hypochemical protein.
SQ SEQUENCE 241 AA; 26719 MW; BE29EE4FEFEE6AB CRC64;
Query Match 17.8%; Score 73.5; DB 4; Length 241;
Best Local Similarity 23.0%; Pred. No. 0.95;
Matches 17; Conservative 20; Mismatches 32; Indels 5; Gaps 2;
QY 8 CMLCRALIKRIQMIIPKALRV-----AVAQVCRVPLVAGICOCCLAERYSVILDTLLG 63
```

Db	26	CEVCNNVVQKLDHMLMSNSBELMITHALERVCSVMPASITKECIIIVDTYSPSLVQ-LVA	84
OY	64	RMLPOLVGRVLVRC	77
	:	:	:
Db	85	KITPEKVCCKFIRLC	98

Search completed: December 3, 2003, 15:52:32  
Job time : 57.4286 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 15:40:35 / Search time 32.381 Seconds  
(without alignments)  
171.565 Million cell updates/sec

Title: US-09-788-308D-3

Perfect score: 171  
Sequence: 1 FGIPCCPMLKRLIVVVVVVIVVIGALMGL 35

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

A\_Geneseq\_19Jun03.\*  
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	171	100.0	35	15	AAV14296
2	171	100.0	35	15	AAAR62821
3	171	100.0	35	16	AAAR75432
4	171	100.0	35	17	AAAR97275
5	171	100.0	35	22	AAAB31570
6	171	100.0	35	22	AAU07652
7	171	100.0	38	12	AAAR15607
8	171	100.0	53	9	AAAR18889
9	171	100.0	79	10	AAAP90037

10	171	100.0	79	11	AAAR06333
11	171	100.0	197	8	AAAP70439
12	171	100.0	197	8	AAAP70440
13	171	100.0	197	8	AAAR82935
14	171	100.0	197	9	AAAR80652
15	171	100.0	197	9	AAAR82977
16	171	100.0	197	9	AAAR82978
17	171	100.0	197	10	AAAP90038
18	171	100.0	197	12	AAAR15608
19	171	100.0	197	12	AAAR15608
20	171	100.0	197	24	AAU08968
21	171	100.0	216	9	AAAB58144
22	165	96.5	41	9	AAAP80580
23	165	96.5	60	9	AAAP80578
24	165	96.5	60	9	AAAP80633
25	165	96.5	187	12	AAAR15612
26	165	96.5	250	10	AAAP2068
27	165	96.5	251	11	AAAR05419
28	164	95.9	35	11	AAAR04845
29	162	94.7	187	10	AAAP8407
30	162	94.7	251	12	AAAR15611
31	156.5	91.5	50	12	AAAR15606
32	156.5	91.5	196	10	AAAP1879
33	151	88.3	36	10	AAAP90054
34	151	88.3	36	12	AAAR15603
35	150	87.7	35	16	AAAR15603
36	150	87.7	35	17	AAAR7277
37	148	86.5	32	16	AAAR75435
38	148	86.5	34	16	AAAR75435
39	148	86.5	34	17	AAAR7276
40	145	84.8	35	22	AAAB31582
41	143	83.6	35	17	AAAR15612
42	143	83.6	35	22	AAAB31572
43	139	81.3	34	22	AAAB31583
44	139	81.3	186	17	AAAR03381
45	137	80.1	35	22	AAAB46447

#### ALIGNMENTS

RESULT 1	
AAV14296	AAV14296 standard; peptide; 35 AA
AAV14296	
AC	AAV14296;
XX	
XX	29-JUN-1999 (first entry)
DT	
XX	
XX	Lung surfactant peptide.
DE	
XX	
XX	Lung surfactant; viral disease; therapy; respiratory tract; glycoprotein;
KW	outer envelope; influenza; parainfluenza; respiratory syncytial virus;
KW	measles virus; mumps virus.
KW	
XX	
OS	Synthetic.
XX	
PN	W09400131-A1.
XX	
PD	06-JAN-1994.
XX	
PF	23-JUN-1993; 93WO-JP00851.
XX	
PR	24-JUN-1992; 92JP-0165875.
XX	
PA	(TANB ) TOKYO TANABE CO LTD.
XX	(TANB ) TOKYO TANABE CO.
XX	
PI	Kido H, Sakai K, Sekido S, Tashiro M,
XX	
DR	WPI; 1994-025874/03.
XX	
PT	Agent contg. lung surfactant for treating viral respiratory

Human alveolar surfactant protein A (SP-A) sequence of a canis lupus familiaris (dog) SAP (Val) - synthe Deduced sequence o Human SP5 protein. Human SP5 protein. Deduced sequence o SP-5 clone #19. H SP-5 clone #19. H Human wild-type su Lung cancer associ Human SAP (Val) (1-41 Human SAP (Val) (1-60 Synthetic hydropho SP-C from PC210SP-Fusion protein com CAT:SP-C hybrid pr Alveolar protein. Sequence of surfac SP-C from PC210SP-ASP-5 analogue (4) Deduced sequence o Human derived pept ASP-5 analogue (1) Synthetic peptide Proline surfactant Synthetic peptide Synthetic peptide Bovine surfactant Amino acid fragmen Analogue of lung s Amino acid fragmen Amino acid fragmen Lung surfactant pr Human surfactant p

PT infections - useful against virus with glyco-protein coat. esp.  
PT influenza, parainfluenza, measles, mumps and respiratory syncytial  
PT virus

PS Claim 13; Page 24; 29pp; Japanese.

CC This sequence represents a lung surfactant peptide, used in the agent of  
CC the invention. The agent is for the prevention and treatment of viral  
CC diseases of the respiratory tract contains a lung surfactant. The agent  
CC is used for viruses with glycoprotein in the outer envelope, especially  
CC influenza, parainfluenza, respiratory syncytial, measles and mumps virus.  
CC It acts by attacking the glycoprotein, making the virus more susceptible  
CC to the activity of the cells in the mucous membrane of the upper  
CC respiratory tract and so reducing the severity and inflammation of any  
CC infection. The agent is administered as an aerosol.

XX Sequence 35 AA;

Query Match 100.0%; Score 171; DB 15; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.2e-15;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FGIPCCPVHLKRLIIVVVVLIIVVIVGALLMGL 35  
DB 1 FGIPCCPVHLKRLIIVVVVLIIVVIVGALLMGL 35

## RESULT 2

AAR62821  
ID AAR62821 standard; peptide; 35 AA.

AC AAR62821;

DT 25-MAR-2003 (updated)  
DT 14-JUL-1995 (first entry)

DE Hydrophobic, lung surfactant peptide SP-C.

KW Hydrophobic polypeptide; purification; lung surfactant peptide;  
KW surfactant apoprotein C.

OS Synthetic.

PN MO9425480-A1.

PD 10-NOV-1994.

PF 28-APR-1994; 94WO-JP00731.

PR 30-APR-1993; 93JP-0103957.

PA (TANB ) TOKYO TANABE CO.

PI Aiba T, Fujiwara T, Sakai K, Takei T;

DR WPI; 1994-358185/44.

PT Purification of hydrophobic lung surfactant peptide(s) - by HPLC  
PT using a mixed solvent contg. 3-10% tri:fluoro:acetic acid, and a  
PT packing contg. polyvinyl alcohol

PS Claim 5; Page 13; 19pp; Japanese.

CC This is one of 11 preferred hydrophobic peptides which can be  
CC purified by HPLC using a mixed solvent system contg. 3-10% (v/v)  
CC tri:fluoro:acetic acid and polyvinyl alcohol as column packing. The  
CC hydrophobic peptides purified by this novel method have stronger  
CC surfactant activity than those prepared by conventional methods.  
CC The peptides are useful in lung surfactant formulations.  
CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 35 AA;

Query Match 100.0%; Score 171; DB 15; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.2e-15;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FGIPCCPVHLKRLIIVVVVLIIVVIVGALLMGL 35  
DB 1 FGIPCCPVHLKRLIIVVVVLIIVVIVGALLMGL 35

## RESULT 3

AAR75432  
ID AAR75432 standard; peptide; 35 AA.

AC AAR75432;

DT 01-FEB-1996 (first entry)

DE Synthetic peptide used in a lung surfactant.

KW Lung surfactant; respiratory distress syndrome; hydrophobic; peptide.

OS Synthetic.

PN MO9515980-A1.

PD 15-JUN-1995.

PF 07-DEC-1994; 94WO-JP02057.

PR 08-DEC-1993; 93JP-0307657.

PA (TANB ) TOKYO TANABE CO.

PI Ohkawa H, Ohtsubo E, Takei T;

DR WPI; 1995-224289/29.

PT Lung surfactant containing new synthetic peptide - having an  
PT hydrophobic C-terminal chain, for treatment of respiratory distress  
PT syndrome

PS Disclosure; Page 40; 54pp; Japanese.

CC AAR75432-R75452 are synthetic peptides with hydrophobic C-terminal  
CC chains. The peptides can be formulated with a lipid mixture  
CC (choline phosphoglyceride, acidic phospholipid and a fatty acid)  
CC to give a lung surfactant. The surfactant is used to treat  
CC respiratory stress disorder.

XX Sequence 35 AA;

Query Match 100.0%; Score 171; DB 16; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.2e-15;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FGIPCCPVHLKRLIIVVVVLIIVVIVGALLMGL 35  
DB 1 FGIPCCPVHLKRLIIVVVVLIIVVIVGALLMGL 35

## RESULT 4

AAR97275  
ID AAR97275 standard; peptide; 35 AA.

AC AAR97275;

DT 11-FEB-1997 (first entry)

DE Human surfactant peptide C, SP-C.

KW Surfactant; respiratory distress syndrome; intermediate; soluble;  
KW treatment; human; surfactant peptide C; SP-C.

XX

OS Homo sapiens.  
XX WO9617872-A1.  
PN 13-JUN-1996.  
XX  
XX 06-JUN-1995; 95WO-JP01114.  
XX 07-DEC-1994; 94JP-0303397.  
XX (TANB ) TOKYO TANABE CO.  
XX Ohtsubo E, Takei T;  
XX WPI; 1996-287121/29.  
XX  
XX Peptide intermediate for production of surfactant peptide(s) - used  
PT in lung surfactants for treatment of respiratory distress syndrome  
XX  
XX Disclosure; Page 11; 19pp; Japanese.  
XX  
XX The present sequence is the human surfactant peptide C (SP-C), upon  
CC which the SP intermediates of the invention are based. SP prep.  
CC from the intermediates may be incorporated into lung surfactant  
CC formulations for the treatment of respiratory distress syndrome,  
CC and are highly soluble in, e.g. methanol, are readily  
CC compounded with lipid mixers and have good suspendability and  
CC surfactant activities.  
XX  
SQ Sequence 35 AA;  
Query Match 100.0%; Score 171; DB 17; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1,2e-15;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FGIPCCPVHLKRLIIYVVVVVLIYVYVGLMLGL 35  
DB 1 FGIPCCPVHLKRLIIYVVVVVLIYVYVGLMLGL 35  
RESULT 5  
AAB31570  
ID AAB31570 standard; peptide; 35 AA.  
XX  
XX AAB31570;  
XX  
XX 20-APR-2001 (first entry)  
XX  
XX Amino acid sequence of the central part of human protein SP-C.  
XX  
XX Surfactant protein C; SP-C; pneumonia; bronchitis; cystic fibrosis;  
XX meconium aspiration syndrome; chronic obstructive pulmonary disease;  
XX COPD; asthma; infant respiratory distress syndrome; IRDS;  
XX acute lung injury; ALI; adult respiratory distress syndrome; ARDS.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX Misc-difference 1 /note= "optionally absent"  
XX  
XX WO200078810-A1.  
XX  
XX 28-DEC-2000.  
XX  
XX 02-JUN-2000; 2000WO-EP05031.  
XX  
XX 17-JUN-1999; 99EP-0111728.  
XX  
XX (BYKG ) BYK GULDEN LOWBERG CHEM FAB.  
XX  
XX Ise W, Gerhardt W, Haefner D, Ulrich W, Sturm E;  
XX  
XX

DR WPI; 2001-091552/10.  
XX  
XX Novel surfactant protein C esters useful for preparing pharmaceutical  
PT compositions for treating infant and adult respiratory distress  
PT syndrome, pneumonia, asthma, cystic fibrosis -  
XX  
XX Disclosure; Page 19; 22pp; English;  
XX  
XX The present sequence represents the central part of a human surfactant  
CC protein C (SP-C). The specification describes SP-C proteins where the  
CC amino acid at the carboxy terminus is esterified with an alcohol having  
CC 1-4 carbon atoms, or its acceptable salts. The SP-C esters have a low  
CC tendency to aggregate and have improved stability in solution.  
CC Pharmaceutical compositions comprising esterified SP-C are useful for  
CC treatment and prophylaxis of pneumonia, bronchitis, meconium  
CC aspiration syndrome, chronic obstructive pulmonary disease (COPD),  
CC asthma and cystic fibrosis, infant respiratory distress syndrome (IRDS)  
CC and/or acute lung injury (ALI) including acute or adult respiratory  
CC distress syndrome (ARDS).  
XX  
SQ Sequence 35 AA;  
Query Match 100.0%; Score 171; DB 22; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1,2e-15;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 FGIPCCPVHLKRLIIYVVVVVLIYVYVGLMLGL 35  
DB 1 FGIPCCPVHLKRLIIYVVVVVLIYVYVGLMLGL 35  
RESULT 6  
AAU07652  
ID AAU07652 standard; Protein; 35 AA.  
XX  
XX AAU07652;  
XX  
XX 04-DEC-2001 (first entry)  
XX  
XX Human surfactant protein C (SP-C).  
XX  
XX Human; surfactant protein B; surfactant protein C; SP-B; SP-C; peptoid;  
XX pulmonary; protein therapy; spreading agent; N-substituted glycine;  
XX lung surfactant; pulmonary surfactant; alveolar surface activity;  
XX respiratory distress syndrome.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX Modified-site 5..6 /note= "palmitoylated"  
XX  
XX WO200160837-A2.  
XX  
XX 23-AUG-2001.  
XX  
XX 16-FEB-2001; 2001WO-US05145.  
XX  
XX 16-FEB-2000; 2000US-0182847.  
XX  
XX (NOUN ) UNIV NORTHWESTERN.  
XX (CHIR ) CHIRON CORP.  
XX  
XX Barron AE, Zuckermann RN, Wu CW;  
XX  
XX WPI; 2001-550045/61.  
XX  
XX Heteropolymetric pulmonary spreading agent having at least one  
PT N-substituted glycine residue and an amino acid residue corresponding  
PT to a natural surfactant-associated protein, useful for treating lung  
PT respiratory distress -  
XX  
XX Claim 1; Fig 5; 40pp; English.  
XX  
XX

XX The invention relates to a non-natural heteropolymetric pulmonary  
CC spreading agent comprising at least one N-substituted glycine residue and  
CC at least one amino acid residue corresponding to the surfactant proteins  
CC B and C (SP-B and SP-C). These surfactant protein mimics are peptid  
CC sequences (reverse sequence of the natural protein i.e. equal to the  
CC carboxy to amino sequence of the peptide) added to a lipid mixture to  
CC create a functional, non-immunogenic lung surfactant with physiological  
CC alveolar surface activity. The peptid sequences of the invention can be  
CC used to enhance the solubility of surfactant associated proteins (to  
CC therefore enhance resistance to aggregation) and can also affect alveolar  
CC surface tension during an inhalation/exhalation cycle. The spreading  
CC agents are useful for treating disorders of the lungs such as respiratory  
CC distress syndrome. This sequence represents the human surfactant protein  
CC C (SP-C).  
XX  
SQ Sequence 35 AA;  
Query Match 100.0%; Score 171; DB 22; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.2e-15;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 FGIPCCPVHLKRLIIVVVVLIIVVIVGALLMGL 35  
Db 1 FGIPCCPVHLKRLIIVVVVLIIVVIVGALLMGL 35  
RESULT 7  
AAR15607  
ID AAR15607 standard; Protein; 38 AA.  
XX  
AC AAR15607;  
XX  
DT 25-MAR-2003 (updated)  
DT 16-MAR-1992 (first entry)  
XX  
DE ASP-5 analogue (5).  
XX  
KM Alveolar surfactant protein; aggregation; extraction; purification.  
XX  
OS Homo sapiens.  
XX  
PN WO9118015-A.  
XX  
PD 28-NOV-1991.  
XX  
PF 17-MAY-1991; 91WO-US03490.  
XX  
PR 17-MAY-1990; 90US-0524360.  
XX  
PA (CALD ) CALIFORNIA BIOTECHNOLOGY INC.  
XX  
PI Benson BJ, White RT, Schilling JW, Buckley DI, Scarborough RM;  
XX WPI; 1991-369185/50.  
XX  
DR New alveolar surfactant protein analogues - used for treating  
XX PT respiratory distress syndrome, pneumonia and bronchitis  
XX  
PS Claim 8; Page 43; 67pp; English.  
XX  
CC The polypeptide may be prepd. by solid phase peptide synthesis or by  
CC using recombinant DNA techniques. The ASP analogue, in addition to  
CC retaining the stability and biological activity of the native  
CC polypeptide, is less susceptible to aggregation than the native  
CC polypeptide. The analogue is therefore much easier to extract and  
CC purify than the native polypeptide.  
CC (See also AAR15602-7; AAQ15262-63 and AAQ15265-66.  
CC (Updated on 25-MAR-2003 to correct PA field.)  
XX  
SQ Sequence 38 AA;  
Query Match 100.0%; Score 171; DB 12; Length 38;

Best Local Similarity 100.0%; Pred. No. 1.3e-15;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 FGIPCCPVHLKRLIIVVVVLIIVVIVGALLMGL 35  
Db 1 FGIPCCPVHLKRLIIVVVVLIIVVIVGALLMGL 35  
RESULT 8  
AAP81889  
ID AAP81889 standard; protein; 53 AA.  
XX  
AC AAP81889;  
XX  
DT 25-MAR-2003 (updated)  
DT 03-FEB-1991 (first entry)  
XX  
DE Sequence encoded by human pulmonary hydrophobic surfactant-associated  
DE protein (SAP) (Val) genomic clone exon 2.  
XX  
KM Human pulmonary hydrophobic surfactant-associated protein (SAP) (Val);  
KM hyaline membrane disease (HMD); therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO8803170-A.  
XX  
PD 05-MAY-1988.  
XX  
PF 02-OCT-1987; 87WO-US02536.  
XX  
PR 01-OCT-1987; 87US-0101680.  
PR 08-DEC-1986; 86US-0939206.  
PR 10-JUN-1987; 87US-0060719.  
XX  
PA (WHIT/) WHITSETT J A.  
PA (ABBO ) ABBOTT LAB.  
XX  
PI Whitsett JA, Fox JL, Pilotmacia TJ, Meuth JL, Sarin VK;  
XX WPI; 1988-133244/19.  
XX  
DR N-PSDB; AAN80644.  
XX  
PT Pulmonary hydrophobic surfactant-associated proteins -  
PT used with lipid(s) to treat and prevent hyaline membrane disease  
PT and similar syndromes  
XX  
PS Example; Fig 10a-10e; p144; English.  
XX  
CC The sequence is derived from human embryonic kidney cell genomic  
CC library. SAP (Val) and SAP (Phe), when combined with lipids, have  
CC significant pulmonary biophysical surfactant activity that may be  
CC utilized to treat and prevent hyaline membrane disease (HMD) and  
CC other syndromes associated with lack or insufficient amts. of natural  
CC pulmonary surfactant material. Antibodies and antisera may also be  
CC made which are directed against SAP (Val) or SAP (Phe). SAP (Val) and  
CC SAP (Phe) in body fluids may be assayed using the compns.  
CC (Updated on 25-MAR-2003 to correct PI field.)  
XX  
SQ Sequence 53 AA;  
Query Match 100.0%; Score 171; DB 9; Length 53;  
Best Local Similarity 100.0%; Pred. No. 1.9e-15;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 FGIPCCPVHLKRLIIVVVVLIIVVIVGALLMGL 35  
Db 10 FGIPCCPVHLKRLIIVVVVLIIVVIVGALLMGL 44  
RESULT 9  
AAP90037

```

ID AAP90037 standard; protein; 79 AA.
XX
XX AAP90037;
AC
XX
DT 25-MAR-2003 (updated)
DT 10-APR-1990 (first entry)
XX
DE Sequence encoded in human SP-5 cDNA.
XX
XX Alveolar surfactant protein activity; ASP activity;
KM respiratory distress syndrome; RDS
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Peptide 24..74
FT Peptide 34..74
FT Peptide 24..61
FT Peptide 31..61
FT Peptide 30..61
FT Peptide 28..61
FT Peptide 26..61
FT Peptide 24..61
XX
XX MO8904326-A.
XX
XX 18-MAY-1989.
XX
XX 02-NOV-1988; 88MO-US03899.
XX
XX 04-NOV-1987; 87US-0117099.
XX
XX 01-NOV-1988; 88US-0266443.
XX
XX (CALD ) CALIFORNIA BIOTECHNOLOGY INC.
XX
XX Benson BJ, White RT, Schilling JW, Buckley D, Scarborough RM;
PI WPI; 1989-165617/22.
XX
XX N-PSDB; AAN90095.
DR
XX
XX Human SP-18 and SP-5 derived peptide(s)
PT - with alveolar surfactant protein activity, used for
PT treating respiratory distress syndrome, pneumonia and
PT bronchitis
XX
XX Claim 6; Figure 3; 63pp; English.
PS
XX
XX Various synthetic peptides based on AAP90037 have been synthesised using
CC standard techniques and are believed by the inventors to be useful as
CC alveolar surfactants in treating respiratory distress syndrome. Some of
CC these peptides are specifically claimed (Claim 6).
CC (Updated on 25-MAR-2003 to correct PR field.)
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 79 AA;
SQ
XX
XX Query Match 100.0%; Score 171; DB 10; Length 79;
XX Best Local Similarity 100.0%; Pred. No. 2.8e-15;
XX Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35
QY |||||
DB 24 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 58
|||

RESULT 10
AAR06333
ID AAR06333 standard; protein; 79 AA.
XX
XX AAR06333;
AC
XX
DT 07-DEC-1990 (first entry)
XX

```

```

DE Human alveolar surfactant protein (SP-C).
XX
XX Alveolar surfactant; SP-A; SP-B; SP-C; respiratory distress syndrome;
KM oxygen toxicity; alpha-1-anti-protease; emphysema; lung cancer;
KM bronchitis; asthma; tuberculosis;
XX
XX Homo sapiens.
OS
XX
XX MO8007469-A.
XX
XX 12-JUL-1990.
XX
XX 29-DEC-1989; 89MO-US00587.
XX
XX 29-DEC-1988; 88US-0295926.
XX
XX (BENS/) BENSON B J.
XX
XX Benson BJ, Wright J;
PI WPI; 1990-238980/31.
XX
XX Pulmonary admin. of liposome contg. active cpds. - uses alveolar
PT surfactant protein to enhance transport across lung surface
PT useful for treating variety of lung specific diseases
XX
XX Disclosure; Page 7; 7pp; English.
XX
XX Alveolar surfactant protein is useful in enhancing the uptake of
CC liposomes containing a pharmacuetically active compound (local or
CC systemic) across the pulmonary surface.
CC This is useful in treating a variety of lung specific diseases
CC eg. respiratory distress syndromes, pneumonia, oxygen toxicity,
CC alpha-1-anti-protease deficiency, emphysema; asthma, tuberculosis,
CC lung cancer and bronchitis.
XX
XX Sequence 79 AA;
SQ
XX
XX Query Match 100.0%; Score 171; DB 11; Length 79;
XX Best Local Similarity 100.0%; Pred. No. 2.8e-15;
XX Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35
QY |||||
DB 24 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 58
|||

RESULT 11
AAP70439
ID AAP70439 standard; protein; 197 AA.
XX
XX AAP70439;
AC
XX
XX 25-MAR-2003 (updated)
DT 17-JAN-1991 (first entry)
XX
XX Sequence of a canine 5 kd alveolar surfactant protein (ASP)
DE from clone cDNA #18.
XX
XX Lung surfactant; respiratory disease syndrome; therapy.
XX
XX Dog.
XX
XX MO8706588-A.
XX
XX 05-NOV-1987.
XX
XX 30-APR-1987; 87MO-US00978.
XX
XX 30-APR-1986; 86US-0857715.
XX
XX 29-JAN-1987; 87US-0008453.
XX
XX (CALD ) CALIFORNIA BIOTECHNOLOGY INC.
XX

```

XX Schilling JW;  
PI White RT;  
PI Cordell B;  
PI Benson BJ;  
XX WPI; 1987-320974/45.  
DR N-PSDB; AAN70714.  
XX  
XX Pure alveolar surfactant protein - obtd. by recombinant DNA methods  
PT and affinity chromatography for treating respiratory disease  
PT syndrome.  
XX  
XX Claim 2; Fig 5; 73pp; English.  
XX  
XX An ASP encoded by human SP-18 DNA and human SP-5 DNA is claimed.  
CC The purified ASP is suitable for treating respiratory disease  
CC syndrome in mammals, esp. when administered with a phospholipid  
CC and opt. with the 32k ASP protein. cDNAs #s 18 and 19 differ by  
CC four nucleotides.  
CC (Updated on 25-MAR-2003 to correct PR field.)  
CC (Updated on 25-MAR-2003 to correct PA field.)  
XX  
SQ Sequence 197 AA;  
XX  
Query Match 100.0%; Score 171; DB 8; Length 197;  
Best Local Similarity 100.0%; Pred. No. 7e-15;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 FGIPCCPVHLKRLIIVVVVLIIVVIGALLMGL 35  
DB 24 FGIPCCPVHLKRLIIVVVVLIIVVIGALLMGL 58  
XX  
RESULT 12  
AAP70440  
ID AAP70440 standard; protein; 197 AA.  
XX  
AC AAP70440;  
XX  
DT 25-MAR-2003 (updated)  
DT 17-JAN-1991 (first entry)  
XX  
DE Sequence of a canine 5 kd alveolar surfactant protein (ASP)  
DE from clone cDNA #19.  
XX  
KM Lung surfactant; respiratory disease syndrome; therapy.  
XX  
OS Dog.  
XX  
PN WO8706588-A.  
XX  
PD 05-NOV-1987.  
XX  
PF 30-APR-1987; 87WO-US00978.  
XX  
PR 30-APR-1986; 86US-0857715.  
PR 29-JAN-1987; 87US-0008453.  
XX  
XX (CALD ) CALIFORNIA BIOTECHNOLOGY INC.  
PA  
XX Schilling JW;  
PI White RT;  
PI Cordell B;  
PI Benson BJ;  
XX  
DR WPI; 1987-320974/45.  
DR N-PSDB; AAN70715.  
XX  
PT Pure alveolar surfactant protein - obtd. by recombinant DNA methods  
PT and affinity chromatography for treating respiratory disease  
PT syndrome.  
XX

PS Claim 2; Fig 6; 73pp; English.  
XX  
XX An ASP encoded by human SP-18 DNA and human SP-5 DNA is claimed.  
CC The purified ASP is suitable for treating respiratory disease  
CC syndrome in mammals, esp. when administered with a phospholipid  
CC and opt. with the 32k ASP protein. cDNAs #s 18 and 19 differ by  
CC four nucleotides.  
CC (Updated on 25-MAR-2003 to correct PR field.)  
CC (Updated on 25-MAR-2003 to correct PA field.)  
XX  
SQ Sequence 197 AA;  
XX  
Query Match 100.0%; Score 171; DB 8; Length 197;  
Best Local Similarity 100.0%; Pred. No. 7e-15;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 FGIPCCPVHLKRLIIVVVVLIIVVIGALLMGL 35  
DB 24 FGIPCCPVHLKRLIIVVVVLIIVVIGALLMGL 58  
XX  
RESULT 13  
AAP82935  
ID AAP82935 standard; protein; 197 AA.  
XX  
AC AAP82935;  
XX  
DT 25-MAR-2003 (updated)  
DT 21-NOV-1990 (first entry)  
XX  
DE SAP (Val).  
XX  
KM SAP (Val); SAP (Phe); hyaline membrane disease.  
XX  
OS synthetic.  
XX  
PN WO8804324-A.  
XX  
PD 16-JUN-1988.  
XX  
PF 03-DEC-1987; 87WO-US03180.  
XX  
PR 08-DEC-1986; 86US-0939206.  
PR 10-JUN-1987; 87US-0060719.  
PR 01-OCT-1987; 87US-0101680.  
XX  
PA (ABBO ) ABBOTT LABORATORIES.  
XX  
PI Whitsett JA, Fox JL, Pilotmatia TJ, Neuth JL;  
XX  
DR WPI; 1988-175472/25.  
DR N-PSDB; AAN80617.  
XX  
PT Pulmonary hydrophobic surfactant-associated proteins - useful for  
PT normalising pulmonary surface tension.  
XX  
PS Disclosure; Page 7; 7pp; English.  
XX  
XX This pulmonary hydrophobic surfactant-associated protein SAP(Val)  
CC is encoded by a genomic clone following processing to remove  
CC introns. When SAP(Phe) is combined with a lipid it can be used to  
CC reduce or maintain normal pulmonary surface tension in the alveoli  
CC of animals (esp. humans) and therefore can be used in the treatment  
CC of hyaline membrane disease in premature infants. It may also be  
CC used to deliver substances to respiratory epithelial cells.  
CC This protein can also be encoded by a cDNA clone which does not  
CC contain the introns.  
CC See also AAP80572-96, AAN80615-16 and AAN82412-13.  
CC (Updated on 25-MAR-2003 to correct PR field.)  
CC (Updated on 25-MAR-2003 to correct PI field.)  
XX  
SQ Sequence 197 AA;  
XX



Query Match 100.0%; Score 171; DB 9; Length 197;  
Best Local Similarity 100.0%; Pred. No. 7e-15;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35  
Db 24 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 58

## RESULT 14

AAP80652 ID AAP80652 standard; protein; 197 AA.

XX AAP80652;

DT 25-MAR-2003 (updated)  
DT 17-SEP-1990 (first entry)

DE Deduced sequence of human pulmonary hydrophobic surfactant-associated protein (SAP) (Val).

KM Human pulmonary hydrophobic surfactant-associated protein (SAP) (Val);  
hyaline membrane disease (HMD) prevention; assays.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Region 25..42 /note="sequence obtained directly on human SAP(Val) protein"

FT Region 39..57 /note="hydrophobic"

PN MO803170-A.

XX 05-MAY-1988.

XX 02-OCT-1987; 87WO-US02536.

XX 08-DEC-1986; 86US-0939206.

PR 10-JUN-1987; 87US-0060719.

PR 01-OCT-1987; 87US-0101680.

XX (WHIT/) WHITSETT J A.

PA (ABBO) ABBOTT LAB.

XX Whitehead JA, Fox JL, Pilotmatia TJ, Meuth JL, Sarin VK;

XX WPI; 1988-133244/19.

DR N-PSDB; AAN80641.

XX Pulmonary hydrophobic surfactant-associated proteins -  
used with lipid(s) to treat and prevent hyaline membrane disease  
and similar syndromes

XX Example; Fig 5a-5b; 144pp; English.

XX It is the deduced sequence of overlapping cDNA clones 311.3, 13-1,  
7T11-2, TP9-1 and R22-1. The predicted sequence for residues 25-42  
(see FT) obtained sequence match at 16 or 17 AAs, the difference being  
His32 instead of Asn. The hydrophobic regions of SAP (Phe) (AAP80651)  
and SAP (Val) (AAP80652) are somewhat homologous. Although these two  
proteins are encoded by distinct genes, it is believed that they are  
structurally related. SAP (Val) and SAP (Phe), when combined with  
lipid, have significant pulmonary biophysical surfactant activity that  
may be utilized to treat and prevent hyaline membrane disease (HMD) and  
other syndromes associated with lack or insufficient amts. of natural  
pulmonary surfactant material. Antibodies and antisera may also be made  
which are directed against SAP (Val) or SAP (Phe). SAP (Val) and SAP  
(Phe) in body fluids may be assayed using the compns.

XX (Updated on 25-MAR-2003 to correct PR field.)

XX (Updated on 25-MAR-2003 to correct PA field.)

XX (Updated on 25-MAR-2003 to correct PI field.)

XX (Updated on 25-MAR-2003 to correct PI field.)

XX Sequence 197 AA;

Qy 1 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35  
Db 24 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 58

Qy 1 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35  
Db 24 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 58

## RESULT 15

AAP82977 ID AAP82977 standard; protein; 197 AA.

XX AAP82977;

DT 25-MAR-2003 (updated)  
DT 04-DEC-1990 (first entry)

DE Human SPS protein.

KM Alveolar surfactant protein; ASP; respiratory distress syndrome;  
pneumonia; bronchitis; SPS.

XX Homo sapiens.

XX MO805820-A.

XX 11-AUG-1988.

XX 15-JAN-1987; 87WO-US00092.

XX 30-APR-1986; 86US-0857715.

PR 29-JAN-1987; 87US-0008453.

XX (CALB-) CALIF BIOTECHN INC.

XX Schilling JW, White RT, Cordell B, Benson BJ;

XX WPI; 1988-124493/33.

DR N-PSDB; AAN80720.

XX Recombinant alveolar surfactant protein - used for treating  
respiratory distress syndrome and related diseases e.g. pneumonia  
and bronchitis.

XX Disclosure; Page 7; ?pp; English.

XX The sequence was deduced from DNA carried on clone No. 18,  
isolated from a human lung cDNA library in lambda gt10 using  
probes prep. from the corresp. canine sequence. The protein  
is the alveolar surfactant protein SPS, of putative mol. wt 19 kd  
(processed to the 5 or 8 kd proteins found in extracts). The SPS  
protein is a member of the low mol. wt., hydrophobic 10k ASP gp.

XX The DNA sequence can be expressed as a recombinant protein and  
used for the treatment of respiratory disorders.

XX See also AAP82978-80, AAP82982 and AAP80694.

XX (Updated on 25-MAR-2003 to correct PR field.)

XX (Updated on 25-MAR-2003 to correct DR field.)

XX Sequence 197 AA;

Qy 1 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35  
Db 24 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 58

Qy 1 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35  
Db 24 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 58

Thu Dec 4 18:32:40 2003

us-09-788-308d-3.rag

Search completed: December 3, 2003, 15:49:54  
Job time : 33.381 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 15:47:35 ; Search time 11.1905 Seconds  
(without alignments)  
132.334 Million cell updates/sec

Title: US-09-788-308D-3  
Sequence: 1 FGIIPCCPVHLKRLIVVVVLIIVVIGALLMGL 35

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2\_6/prodata/1/aa/5A\_COMB.pep.\*  
2: /cgn2\_6/prodata/1/aa/5B\_COMB.pep.\*  
3: /cgn2\_6/prodata/1/aa/5A\_COMB.pep.\*  
4: /cgn2\_6/prodata/1/aa/5B\_COMB.pep.\*  
5: /cgn2\_6/prodata/1/aa/5A\_COMB.pep.\*  
6: /cgn2\_6/prodata/1/aa/5B\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171	100.0	35	1	US-08-343-427B-11
2	171	100.0	35	2	US-08-652-450A-1
3	171	100.0	35	6	5223481-6
4	164	95.9	35	6	5223481-5
5	164	95.9	35	6	5455227-2
6	163	95.3	35	6	5223481-7
7	150	87.7	35	2	US-08-652-450A-3
8	148	86.5	32	2	US-08-652-450A-4
9	148	86.5	34	2	US-08-652-450A-2
10	139	81.3	186	2	US-08-750-194-1
11	137	80.1	35	2	US-08-750-194-2
12	135	78.9	35	2	US-08-652-450A-5
13	129	75.4	27	1	US-08-343-427B-2
14	124	72.5	27	1	US-08-343-427B-1
15	120	70.2	27	1	US-08-343-427B-6
16	118	69.0	27	1	US-08-343-427B-4
17	115	67.3	27	1	US-08-343-427B-5
18	115	67.3	27	1	US-08-343-427B-7
19	115	67.3	27	1	US-08-343-427B-9
20	113	66.1	27	1	US-08-343-427B-3
21	112	65.5	27	2	US-08-652-450A-20
22	111	64.9	27	1	US-08-343-427B-8
23	111	64.9	27	1	US-08-343-427B-10
24	108	63.2	27	2	US-08-652-450A-21
25	88	51.5	27	2	US-08-652-450A-6
26	86	50.3	17	6	5223481-8
27	86	50.3	17	6	5455227-4

28	74	43.3	12	6	5223481-3	Patent No. 5223481
29	74	43.3	12	6	5455227-6	Patent No. 5455227
30	74	43.3	23	2	US-08-652-450A-16	Sequence 16, App1
31	73	42.7	27	2	US-08-652-450A-7	Sequence 7, App1
32	72	42.1	23	2	US-08-652-450A-15	Sequence 15, App1
33	70	40.9	23	2	US-08-652-450A-8	Sequence 8, App1
34	69	40.4	15	3	US-08-445-422-2	Sequence 2, App1
35	69	40.4	15	4	US-09-435-204-2	Sequence 2, App1
36	69	40.4	23	2	US-08-652-450A-17	Sequence 17, App1
37	67	39.2	12	6	5223481-2	Patent No. 5223481
38	67	39.2	12	6	5455227-1	Patent No. 5455227
39	66	38.6	12	6	5223481-4	Patent No. 5223481
40	66	38.6	19	2	US-08-652-450A-9	Sequence 9, App1
41	65	38.0	23	2	US-08-652-450A-18	Sequence 18, App1
42	63	36.8	23	2	US-08-652-450A-10	Sequence 10, App1
43	61	35.7	23	4	US-09-228-986-72	Sequence 72, App1
44	60.5	35.4	910	4	US-09-228-986-72	Sequence 2, App1
45	59	34.5	386	3	US-09-086-483A-2	

#### ALIGNMENTS

RESULT 1  
US-08-343-427B-11  
Sequence 11, Application US/08343427B  
Patent No. 5648457  
GENERAL INFORMATION:  
APPLICANT: TAKEI, Tsunetome; AIBA, Toshimitsu; SAKAI, Kaoru;  
TITLE OF INVENTION: PURIFICATION METHOD FOR HYDROPHOBIC POLYPEPTIDE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McGlew and Tuttle, P.C.  
STREET: Scarborough Station  
CITY: Scarborough  
STATE: New York  
ZIP: 10510-0827  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage  
COMPUTER: ALLUR Technology 386SX (IBM compatible)  
OPERATING SYSTEM: DOS 5.0  
SOFTWARE: Word Perfect 5.1+ for DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,427B  
FILING DATE: 14-NOV-1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP94/00731  
FILING DATE: 28-APR-1994  
APPLICATION NUMBER: Japan 103957/1993  
FILING DATE: 30-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGlew, John J.  
REGISTRATION NUMBER: 17722  
NAME: McGlew, John James  
REGISTRATION NUMBER: 31903  
NAME: McGlew, Hilda S.  
REGISTRATION NUMBER: 30295  
NAME: Dengler, Theobald  
REGISTRATION NUMBER: 34575  
NAME: Ceccon, Clario  
REGISTRATION NUMBER: 19268  
NAME: Goodman, Christopher D.  
REGISTRATION NUMBER: 34338  
REFERENCE/DOCKET NUMBER: 41450  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (914) 941-5600  
TELEFAX: (914) 941-5655  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: Amino acid

TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
US-08-343-427B-11

Query Match 100.0%; Score 171; DB 1; Length 35;  
Best Local Similarity 100.0%; Pred. No. 7.4e-16;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35  
DB 1 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35

RESULT 2  
US-08-652-450A-1  
Sequence 1, Application US/08652450A

Patent No. 5827825  
GENERAL INFORMATION:  
APPLICANT: TAKEI, TSUNETOMO  
APPLICANT: OHTSUBO, EIJI  
APPLICANT: OHKAWA, HIROSHI  
TITLE OF INVENTION: NOVEL SYNTHETIC PEPTIDE, LUNG SURFACTANT  
TITLE OF INVENTION: CONTAINING THE SAME AND REMEDY FOR RESPIRATORY DISTRESS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESS: GRIFPIN, BUTLER, WISEHUNT & KURTOSKY  
STREET: 2300 SOUTH NINTH STREET, SUITE PH-1  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: U.S.A.  
ZIP: 22204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/652,450A  
FILING DATE: 05-JUN-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 307657/1993  
FILING DATE: 08-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SZIDL, JOERG-UWE  
REGISTRATION NUMBER: 31,799  
REFERENCE/DOCKET NUMBER: A0BA0006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 979-5700  
TELEFAX: (703) 979-7429  
INFORMATION FOR SEQ ID NO. 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHEICAL: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Human  
TISSUE TYPE: HUMAN LUNG  
US-08-652-450A-1

Query Match 100.0%; Score 171; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred. No. 7.4e-16;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35  
DB 1 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35

RESULT 3  
5223481-6  
Patent No. 5223481

APPLICANT: CURSTEDT, TORE;JORNVALL, HANS;LOMENADLER, BJORN;  
ROBERTSSON, BENGT  
TITLE OF INVENTION: BIOLOGICALLY ACTIVE LIPOPROTEIN AND  
ITS USE  
NUMBER OF SEQUENCES: 8  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/423,346  
FILING DATE: 18-OCT-1989  
SEQ ID NO.6:  
LENGTH: 35

5223481-6

Query Match 100.0%; Score 171; DB 6; Length 35;  
Best Local Similarity 100.0%; Pred. No. 7.4e-16;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35  
DB 1 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35

RESULT 4  
5223481-5

Patent No. 5223481  
APPLICANT: CURSTEDT, TORE;JORNVALL, HANS;LOMENADLER, BJORN;  
ROBERTSSON, BENGT  
TITLE OF INVENTION: BIOLOGICALLY ACTIVE LIPOPROTEIN AND  
ITS USE  
NUMBER OF SEQUENCES: 8  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/423,346  
FILING DATE: 18-OCT-1989  
SEQ ID NO.5:  
LENGTH: 35

5223481-5

Query Match 95.9%; Score 164; DB 6; Length 35;  
Best Local Similarity 97.1%; Pred. No. 6e-15;  
Matches 34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35  
DB 1 FXIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35

RESULT 5  
5455227-2

Patent No. 5455227  
APPLICANT: CURSTEDT, TORE;LOMENADLER, BJORN;ROBERTSON, BENGT  
TITLE OF INVENTION: BIOLOGICALLY ACTIVE LIPOPROTEIN AND ITS USE  
NUMBER OF SEQUENCES: 6  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/64,382  
FILING DATE: 21-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 423,346  
FILING DATE: 18-OCT-1989  
SEQ ID NO.2:  
LENGTH: 35

5455227-2

Query Match 95.9%; Score 164; DB 6; Length 35;  
Best Local Similarity 97.1%; Pred. No. 6e-15;  
Matches 34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FGIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35  
DB 1 FXIPCCPVHLKRLIIVVVVVLIIVVIVGALLMGL 35

RESULT 6  
5223481-7  
PATENT NO. 5223481  
APPLICANT: CURSTEDT, TORE; JORNVAL, HANS; LOWENADLER, BJORN;  
ROBERTSSON, BENGT  
TITLE OF INVENTION: BIOLOGICALLY ACTIVE LIPOPROTEIN AND  
ITS USE  
NUMBER OF SEQUENCES: 8  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/423,346  
FILING DATE: 18-OCT-1989  
SEQ ID NO: 7  
LENGTH: 35  
5223481-7

Query Match 95.3%; Score 163; DB 6; Length 35;  
Best Local Similarity 97.1%; Pred. No. 8.1e-15;  
Matches 34; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FGIPCCPVHLKRLIVVVVVVIVVIVGALLMGL 35  
DB 1 FRIPCCPVHLKRLIVVVVVVIVVIVGALLMGL 35

RESULT 7  
US-08-652-450A-3  
Sequence 3, Application US/08652450A  
Patent No. 5827825  
GENERAL INFORMATION:  
APPLICANT: TAKEI, TSUNETOMO  
APPLICANT: OHTSUBO, EIJI  
APPLICANT: OKKAWA, HIROSHI  
TITLE OF INVENTION: NOVEL SYNTHETIC PEPTIDE, LUNG SURFACTANT  
TITLE OF INVENTION: CONTAINING THE SAME AND REMEDY FOR RESPIRATORY DISTRESS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: GRIFFIN, BUTLER, WHISENHUNT & KURTOSKY  
STREET: 2300 SOUTH NINTH STREET, SUITE PH-1  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: U.S.A.  
ZIP: 22204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/652,450A  
FILING DATE: 05-JUN-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 307657/1993  
FILING DATE: 08-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SZIPL, JOERG-UWE  
REGISTRATION NUMBER: 31,799  
REFERENCE/DOCKET NUMBER: A0BA0006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 979-5700  
TELEFAX: (703) 979-7429  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:

ORGANISM: PIG  
TISSUE TYPE: PORCINE LUNG  
US-08-652-450A-3

Query Match 87.7%; Score 150; DB 2; Length 35;  
Best Local Similarity 90.9%; Pred. No. 3.9e-13;  
Matches 30; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 3 IPCCPVHLKRLIVVVVVVIVVIVGALLMGL 35  
DB 3 IPCCPVHLKRLIVVVVVVIVVIVGALLMGL 35

## RESULT 8

US-08-652-450A-4  
Sequence 4, Application US/08652450A  
Patent No. 5827825

GENERAL INFORMATION:

APPLICANT: TAKEI, TSUNETOMO

APPLICANT: OHTSUBO, EIJI

APPLICANT: OKKAWA, HIROSHI

TITLE OF INVENTION: NOVEL SYNTHETIC PEPTIDE, LUNG SURFACTANT

TITLE OF INVENTION: CONTAINING THE SAME AND REMEDY FOR RESPIRATORY DISTRESS

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESSES:

ADDRESSEE: GRIFFIN, BUTLER, WHISENHUNT & KURTOSKY

STREET: 2300 SOUTH NINTH STREET, SUITE PH-1

CITY: ARLINGTON

STATE: VA

COUNTRY: U.S.A.

ZIP: 22204

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/652,450A

FILING DATE: 05-JUN-1996

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 307657/1993

FILING DATE: 08-DEC-1993

ATTORNEY/AGENT INFORMATION:

NAME: SZIPL, JOERG-UWE

REGISTRATION NUMBER: 31,799

REFERENCE/DOCKET NUMBER: A0BA0006

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 979-5700

TELEFAX: (703) 979-7429

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 32 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

FRAGMENT TYPE: N-terminal

PUBLICATION INFORMATION:

AUTHORS: Benson, B J

AUTHORS: White, R T

TITLE: Human SP-18 and SP-5 derived peptide(s) -

TITLE: with alveolar surfactant protein activity, used

TITLE: for treating respiratory distress syndrome,

TITLE: pneumonia and bronchitis

JOURNAL: Japanese Patent Publication Hei 3-502095

DATE: 16-May-1991

US-08-652-450A-4

Query Match 86.5%; Score 148; DB 2; Length 32;  
Best Local Similarity 100.0%; Pred. No. 6.5e-13;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 CCGPHLKRLLIVVVVVVIVVIVGALLMGL 35  
Db 1 CCGPHLKRLLIVVVVVVIVVIVGALLMGL 31

RESULT 9  
US-08-652-450A-2

; Sequence 2, Application US/08652450A

; Patent No. 5827825

; GENERAL INFORMATION:

; APPLICANT: TAKEI, TSUNETOMO

; APPLICANT: OHTSUBO, EIJI

; APPLICANT: OKAWA, HIROSHI

; TITLE OF INVENTION: NOVEL SYNTHETIC PEPTIDE, LUNG SURFACTANT

; TITLE OF INVENTION: CONTAINING THE SAME AND REMEDY FOR RESPIRATORY DISTRESS

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: GRIFFIN, BUTLER, WISEHUNT & KURTOSKY

; STREET: 2300 SOUTH NINTH STREET, SUITE PH-1

; CITY: ARLINGTON

; STATE: VA

; COUNTRY: U.S.A.

; ZIP: 22204

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/652,450A

; FILING DATE: 05-JUN-1996

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 307657/1993

; FILING DATE: 08-DEC-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: SZIPL, JOERG-UWE

; REGISTRATION NUMBER: 31,799

; REFERENCE/DOCKET NUMBER: A08A0006

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 979-5700

; TELEFAX: (703) 979-7429

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 34 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; HYPOTHEICAL: NO

; FRAGMENT TYPE: N-terminal

; ORIGINAL SOURCE:

; ORGANISM: COW

; TISSUE TYPE: BOVINE LUNG

; US-08-652-450A-2

Query Match 86.5%; Score 148; DB 2; Length 34;

Best Local Similarity 90.9%; Pred. No. 6.9e-13;

Matches 30; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 3 IPCCPVHLKRLLIVVVVVVIVVIVGALLMGL 35  
Db 2 IPCCPVHLKRLLIVVVVVVIVVIVGALLMGL 34

RESULT 10

US-08-750-194-2

; Sequence 2, Application US/08750194

; Patent No. 5874406

; GENERAL INFORMATION:

; APPLICANT: SCHAFER, Klaus, P.

; TITLE OF INVENTION: Synthetic peptide analogs of lung surface

; TITLE OF INVENTION: Protein Sp-C

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC

; STREET: 400 Seventh Street, N. W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/750,194

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/EP95/02028

; FILING DATE: 27-MAY-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: I. M. Aisenberg

; REGISTRATION NUMBER: 19,007

; REFERENCE/DOCKET NUMBER: 8125/P60707US0

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202/638-6666

; TELEFAX: 202/393-5350

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 186 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-750-194-2

Query Match 81.3%; Score 139; DB 2; Length 186;

Best Local Similarity 91.2%; Pred. No. 5.6e-11;

Matches 31; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 GIPCCPVHLKRLLIVVVVVVIVVIVGALLMGL 35  
Db 153 GIPCCPVHLKRLLIVVVVVVIVVIVGALLMGL 186

RESULT 11

US-08-750-194-1

; Sequence 1, Application US/08750194

; Patent No. 5874406

; GENERAL INFORMATION:

; APPLICANT: SCHAFER, Klaus, P.

; APPLICANT: MELCHERS, Klaus

; TITLE OF INVENTION: Synthetic peptide analogs of lung surface

; TITLE OF INVENTION: Protein Sp-C

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC

; STREET: 400 Seventh Street, N. W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/750,194

; FILING DATE:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/EP95/02028  
FILING DATE: 27-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: I. M. Aisenberg  
REGISTRATION NUMBER: 19,007  
REFERENCE/DOCKET NUMBER: 8125/P60707U50  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202/393-5350  
TELEFAX: 202/393-6666  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: linear  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1  
OTHER INFORMATION: /note= "Xaa is phe or not an amino  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 6  
OTHER INFORMATION: /note= "Xaa is phe or Trp"  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 5  
OTHER INFORMATION: /note= "Xaa is phe or Trp"  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 33  
OTHER INFORMATION: /note= "Xaa is phe or Trp"  
US-08-750-194-1  
Query Match 80.1%; Score 137, DB 2; Length 35;  
Best Local Similarity 91.2%; Pred. No. 1.9e-11;  
Matches 31; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 2 GIPCCPVHLKRLIIVVVVLIIVVIVGALLMGL 35  
Db 2 GIPXPVHLKRLIIVVVVLIIVVIVGALLMGL 35  
RESULT 12  
US-08-652-450A-5  
Sequence 5, Application US/08652450A  
Patent No. 5827825  
GENERAL INFORMATION:  
APPLICANT: TAKEI, TSUNETOMO  
APPLICANT: OHTSUBO, EIJI  
APPLICANT: OHKAWA, HIROSHI  
TITLE OF INVENTION: NOVEL SYNTHETIC PEPTIDE, LUNG SURFACTANT  
TITLE OF INVENTION: CONTAINING THE SAME AND REMEDY FOR RESPIRATORY DISTRESS  
NUMBER OF INVENTIONS: 21  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GRIFFIN, BUTLER, WISENUNT & KUROSSY  
STREET: 2100 SOUTH NINTH STREET, SUITE PH-1  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: U.S.A.  
ZIP: 22204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/652,450A  
FILING DATE: 05-JUN-1996  
CLASSIFICATION: 514

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 307657/1993  
FILING DATE: 08-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: SZIEPL, JOERG-UWE  
REGISTRATION NUMBER: 31,799  
REFERENCE/DOCKET NUMBER: A0BA0006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 979-5700  
TELEFAX: (703) 979-7429  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: N-terminal  
US-08-652-450A-5  
Query Match 78.9%; Score 135, DB 2; Length 35;  
Best Local Similarity 62.9%; Pred. No. 3.5e-11;  
Matches 22; Conservative 13; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 FGIPCCPVHLKRLIIVVVVLIIVVIVGALLMGL 35  
Db 1 FGIPCCPVHLKRLIIVVVVLIIVVIVGALLMGL 35  
RESULT 13  
US-08-343-427B-2  
Sequence 2, Application US/08343427B  
Patent No. 5646457  
GENERAL INFORMATION:  
APPLICANT: TAKEI, Tsunetomo; AIBA, Toshimitsu; SAKAI, Kaoru;  
APPLICANT: FUJIMURA, Tetsuro  
TITLE OF INVENTION: PURIFICATION METHOD FOR HYDROPHOBIC POLYPEPTIDE  
NUMBER OF INVENTIONS: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McGlew and Tuttle, P.C.  
STREET: Scarborough Station  
CITY: Scarborough  
STATE: New York  
ZIP: 10510-0827  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage  
COMPUTER: ALLUR Technology 386SX (IBM compatible)  
OPERATING SYSTEM: DOS 5.0  
SOFTWARE: Word Perfect 5.1+ for DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,427B  
FILING DATE: 14-NOV-1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP94/00731  
FILING DATE: 28-APR-1994  
APPLICATION NUMBER: Japan 103957/1993  
FILING DATE: 30-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGlew, John J.  
REGISTRATION NUMBER: 17722  
NAME: McGlew, John James  
REGISTRATION NUMBER: 31903  
NAME: McGlew, Hilda S.  
REGISTRATION NUMBER: 30295  
NAME: Dengler, Theobald  
REGISTRATION NUMBER: 34575  
NAME: Ceccon, Clario  
REGISTRATION NUMBER: 19268  
NAME: Goodman, Christopher D.  
REGISTRATION NUMBER: 34338  
REFERENCE/DOCKET NUMBER: 41450

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (914) 941-5600  
TELEFAX: (914) 941-5855  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 amino acids  
TYPE: Amino acid  
TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
US-08-343-427B-2

Query Match 75.4%; Score 129; DB 1; Length 27;  
Best Local Similarity 100.0%; Pred. No. 16e-10;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 CCPVHLKRLIIVVVVLIIVVIGAL 31  
DB 1 CCPVHLKRLIIVVVVLIIVVIGAL 27

RESULT 14  
US-08-343-427B-1  
Sequence 1, Application US/08343427B  
Patent No. 5648457  
GENERAL INFORMATION:  
APPLICANT: TAKEI, Tsunetome; AIBA, Toshimitsu; SAKAI, Kaoru;  
APPLICANT: FUJIWARA, Tetsuro  
TITLE OF INVENTION: PURIFICATION METHOD FOR HYDROPHOBIC POLYPEPTIDE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McGlew and Tuttle, P.C.  
STREET: Scarborough Station  
CITY: Scarborough  
STATE: New York  
ZIP: 10510-0827  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage  
COMPUTER: ALLUR Technology 386SX (IBM compatible)  
OPERATING SYSTEM: DOS 5.0  
SOFTWARE: Word Perfect 5.1+ for DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,427B  
FILING DATE: 14-NOV-1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP94/00731  
FILING DATE: 28-APR-1994  
APPLICATION NUMBER: Japan 103957/1993  
FILING DATE: 30-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGlew, John J.  
REGISTRATION NUMBER: 17722  
NAME: McGlew, John James  
REGISTRATION NUMBER: 31903  
NAME: McGlew, Hilda S.  
REGISTRATION NUMBER: 30295  
NAME: Dengler, Theobald  
REGISTRATION NUMBER: 34575  
NAME: Ceccon, Clario  
REGISTRATION NUMBER: 19268  
NAME: Goodman, Christopher D.  
REGISTRATION NUMBER: 34338  
REFERENCE/DOCKET NUMBER: 41450  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (914) 941-5600  
TELEFAX: (914) 941-5855  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 amino acids  
TYPE: Amino acid  
TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
US-08-343-427B-1

Query Match 72.5%; Score 124; DB 1; Length 27;  
Best Local Similarity 100.0%; Pred. No. 7.2e-10;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 CCPVHLKRLIIVVVVLIIVVIGAL 32  
DB 1 CCPVHLKRLIIVVVVLIIVVIGAL 27

RESULT 15  
US-08-343-427B-6  
Sequence 6, Application US/08343427B  
Patent No. 5648457  
GENERAL INFORMATION:  
APPLICANT: TAKEI, Tsunetome; AIBA, Toshimitsu; SAKAI, Kaoru;  
APPLICANT: FUJIWARA, Tetsuro  
TITLE OF INVENTION: PURIFICATION METHOD FOR HYDROPHOBIC POLYPEPTIDE  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McGlew and Tuttle, P.C.  
STREET: Scarborough Station  
CITY: Scarborough  
STATE: New York  
ZIP: 10510-0827  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB storage  
COMPUTER: ALLUR Technology 386SX (IBM compatible)  
OPERATING SYSTEM: DOS 5.0  
SOFTWARE: Word Perfect 5.1+ for DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/343,427B  
FILING DATE: 14-NOV-1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP94/00731  
FILING DATE: 28-APR-1994  
APPLICATION NUMBER: Japan 103957/1993  
FILING DATE: 30-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McGlew, John J.  
REGISTRATION NUMBER: 17722  
NAME: McGlew, John James  
REGISTRATION NUMBER: 31903  
NAME: McGlew, Hilda S.  
REGISTRATION NUMBER: 30295  
NAME: Dengler, Theobald  
REGISTRATION NUMBER: 34575  
NAME: Ceccon, Clario  
REGISTRATION NUMBER: 19268  
NAME: Goodman, Christopher D.  
REGISTRATION NUMBER: 34338  
REFERENCE/DOCKET NUMBER: 41450  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (914) 941-5600  
TELEFAX: (914) 941-5855  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 amino acids  
TYPE: Amino acid  
TOPOLOGY: Linear  
MOLECULE TYPE: Peptide  
US-08-343-427B-6

Query Match 70.2%; Score 120; DB 1; Length 27;  
Best Local Similarity 88.9%; Pred. No. 2.4e-09;  
Matches 24; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 5 CCPVHLKRLIIVVVVLIIVVIGAL 31  
DB 1 CCPVHLKRLIIVVVVLIIVVIGAL 27



Thu Dec 4 18:32:41 2003

us-09-788-308d-3.rat

Search completed: December 3, 2003, 15:54:26  
Job time : 12.1905 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 15:50:00 ; Search time 21.1905 Seconds

(without alignments)  
307.186 Million cell updates/sec

Title: US-09-788-308D-3

Perfect score: 1 FGIPCCPVHLKRLIIVVVVLIIVVIGALLMCL 35

Sequence: 1 FGIPCCPVHLKRLIIVVVVLIIVVIGALLMCL 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCRT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCRTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09C\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10A\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171	100.0	35	11	US-09-788-308D-3
2	171	100.0	197	14	US-10-074-247-2
3	171	100.0	216	9	US-09-925-302-482
4	100	58.5	25	10	US-09-988-842-5
5	68	39.8	25	10	US-09-988-842-24
6	60.5	35.4	910	15	US-10-101-464A-72
7	59	34.5	181	12	US-10-094-749-2274
8	59	34.5	386	12	US-10-137-870-340
9	59	34.5	386	12	US-10-140-021-340
10	59	34.5	386	12	US-10-140-021-340
11	59	34.5	386	12	US-10-140-274-340
12	59	34.5	386	12	US-10-140-471-340
13	59	34.5	386	12	US-10-140-807-340
14	59	34.5	386	12	US-10-140-922-340
15	59	34.5	386	12	US-10-140-924-340

16	59	34.5	386	12	US-10-140-926-340	Sequence 340, App
17	59	34.5	386	12	US-10-141-698-340	Sequence 340, App
18	59	34.5	386	12	US-10-141-702-340	Sequence 340, App
19	59	34.5	386	12	US-10-141-704-340	Sequence 340, App
20	59	34.5	386	12	US-10-142-421-340	Sequence 340, App
21	59	34.5	386	12	US-10-142-432-340	Sequence 340, App
22	59	34.5	386	12	US-10-142-467-340	Sequence 340, App
23	59	34.5	386	12	US-10-143-033-340	Sequence 340, App
24	59	34.5	386	12	US-10-144-994-340	Sequence 340, App
25	59	34.5	386	12	US-10-145-628-340	Sequence 340, App
26	59	34.5	386	12	US-10-145-631-340	Sequence 340, App
27	59	34.5	386	12	US-10-145-633-340	Sequence 340, App
28	59	34.5	386	12	US-10-145-746-340	Sequence 340, App
29	59	34.5	386	12	US-10-145-748-340	Sequence 340, App
30	59	34.5	386	12	US-10-145-823-340	Sequence 340, App
31	59	34.5	386	12	US-10-145-826-340	Sequence 340, App
32	59	34.5	386	12	US-10-145-870-340	Sequence 340, App
33	59	34.5	386	12	US-10-145-876-340	Sequence 340, App
34	59	34.5	386	12	US-10-145-959-340	Sequence 340, App
35	59	34.5	386	12	US-10-146-724-340	Sequence 340, App
36	59	34.5	386	12	US-10-146-725-340	Sequence 340, App
37	59	34.5	386	12	US-10-146-795-340	Sequence 340, App
38	59	34.5	386	12	US-10-147-495-340	Sequence 340, App
39	59	34.5	386	12	US-10-147-501-340	Sequence 340, App
40	59	34.5	386	12	US-10-147-504-340	Sequence 340, App
41	59	34.5	386	12	US-10-147-506-340	Sequence 340, App
42	59	34.5	386	12	US-10-147-509-340	Sequence 340, App
43	59	34.5	386	12	US-10-147-510-340	Sequence 340, App
44	59	34.5	386	12	US-10-147-511-340	Sequence 340, App
45	59	34.5	386	12	US-10-147-529-340	Sequence 340, App

## ALIGNMENTS

RESULT 1  
US-09-788-308D-3  
; Sequence 3, Application US/09788308D  
; Publication No. US20030040468A1  
; GENERAL INFORMATION:  
; APPLICANT: No. US20030040468A1Western University  
; TITLE OF INVENTION: Polypeptide Pulmonary Surfactants  
; FILE REFERENCE: 6374  
; CURRENT APPLICATION NUMBER: US/09/788, 308D  
; CURRENT FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: US 60/182, 847  
; PRIOR FILING DATE: 2000-02-16  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 3  
; LENGTH: 35  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-788-308D-3

Query Match 100.0%; Score 171; DB 11; Length 35;  
Best Local Similarity 100.0%; Pred. No. 3; Indels 0;  
Matches 35; Conservative 0; Mismatches 0;

OR 1 FGIPCCPVHLKRLIIVVVVLIIVVIGALLMCL 35  
DB 1 FGIPCCPVHLKRLIIVVVVLIIVVIGALLMCL 35

RESULT 2  
US-10-074-247-2  
; Sequence 2, Application US/10074247  
; Publication No. US20020197646A1  
; GENERAL INFORMATION:  
; APPLICANT: No. US20020197646A1Lawrence M.  
; APPLICANT: White, Jeffrey A.  
; APPLICANT: Cole, F. Sessions  
; APPLICANT: Hamvas, Aaron

```
; TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated with Interstitial Lung
; FILE REFERENCE: 001107.00229
; CURRENT APPLICATION NUMBER: US/10/074,247
; PRIOR FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: US 60/268,650
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 60/268,991
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 2
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-074-247-2
```

```
Query Match          100.0%; Score 171; DB 14; Length 197;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 FGIPCCPVHLKRLIVVVVVIVVIGALLMGL 35
Db 24 FGIPCCPVHLKRLIVVVVVIVVIGALLMGL 58
```

```
RESULT 3
US-09-925-302-482
; Sequence 482, Application US/09925302
; Patent No. US2002004941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 482
; LENGTH: 216
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (5)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-482
```

```
Query Match          100.0%; Score 171; DB 9; Length 216;
Best Local Similarity 100.0%; Pred. No. 1.8e-13;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 FGIPCCPVHLKRLIVVVVVIVVIGALLMGL 35
Db 49 FGIPCCPVHLKRLIVVVVVIVVIGALLMGL 83
```

```
RESULT 4
US-09-988-842-5
; Sequence 5, Application US/09988842
; Patent No. US20020143105A1
; GENERAL INFORMATION:
; APPLICANT: Johansson, Jan
; TITLE OF INVENTION: DISCORDANT HELIX STABILIZATION FOR PREVENTION
; FILE REFERENCE: 12125-002001
; CURRENT APPLICATION NUMBER: US/09/988,842
```

```
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: US 60/251,662
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: US 60/253,695
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 5
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetically generated peptide
US-09-988-842-5
```

```
Query Match          58.5%; Score 100; DB 10; Length 25;
Best Local Similarity 88.0%; Pred. No. 1.1e-05;
Matches 22; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 9 HLKRLIVVVVVIVVIGALLM 33
Db 1 HLKRLIVVVVVIVVIGALLM 25
```

```
RESULT 5
US-09-988-842-24
; Sequence 24, Application US/09988842
; Patent No. US20020143105A1
; GENERAL INFORMATION:
; APPLICANT: Johansson, Jan
; TITLE OF INVENTION: DISCORDANT HELIX STABILIZATION FOR PREVENTION
; FILE REFERENCE: 12125-002001
; CURRENT APPLICATION NUMBER: US/09/988,842
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: US 60/251,662
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: US 60/253,695
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 24
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetically generated peptide
US-09-988-842-24
```

```
Query Match          39.8%; Score 68; DB 10; Length 25;
Best Local Similarity 48.0%; Pred. No. 0.085;
Matches 12; Conservative 13; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 9 HLKRLIVVVVVIVVIGALLM 33
Db 1 HLKRLIVVVVVIVVIGALLM 25
```

```
RESULT 6
US-10-101-464A-72
; Sequence 72, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
; APPLICANT: Scrabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; FILE REFERENCE: 11000.1020c2
; CURRENT APPLICATION NUMBER: US/10/101,464A
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
```

```

; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 910
; TYPE: PRT
; ORGANISM: Pinus radiata
US-10-101-464A-72
```

```

Query Match
Best Local Similarity 35.4%; Score 60.5; DB 15; Length 910;
Matches 14; Conservative 9; Mismatches 7; Indels 7; Gaps 1;
```

```

Qy 3 IPCCPVH-----LKRLLIVVVVVVLIIVVIGALL 32
Db 660 LPCCVHKKHKSVLNLRKRVILVVVVAIVLCLFLAIL 696
```

```

RESULT 7
US-10-094-749-2274
; Sequence 2274, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2274
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2274
```

```

Query Match
Best Local Similarity 34.5%; Score 59; DB 12; Length 181;
Matches 12; Conservative 4; Mismatches 11; Indels 0; Gaps 0;
```

```

Qy 4 PCCPVHKLRLIVVVVVVLIIVVIGAA 30
Db 115 PCCPVHPOSILVVMCMCAKCVHVCVA 141
```

```

RESULT 8
US-10-137-870-340
```

```

; Sequence 340, Application US/10137870
; Publication No. US20030138883A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C155
; CURRENT APPLICATION NUMBER: US/10/137,870
; PRIOR FILING DATE: 2002-05-03
; PRIOR Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 340
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo Saplen
US-10-137-870-340
```

```

Query Match
Best Local Similarity 34.5%; Score 59; DB 12; Length 386;
Matches 12; Conservative 8; Mismatches 6; Indels 2; Gaps 1;
```

```

Qy 2 GIPCCPVHKLRLIVVVVVVLIIVVIG 29
Db 204 GMLASPHY--LIIIVLVIIILAVVVG 229
```

```

RESULT 9
US-10-140-018-340
; Sequence 340, Application US/10140018
; Publication No. US20030138885A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C158
; CURRENT APPLICATION NUMBER: US/10/140,018
; PRIOR FILING DATE: 2002-05-06
; PRIOR Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 340
; LENGTH: 386
```

```
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-018-340

Query Match          34.5%; Score 59; DB 12; Length 386;
Best Local Similarity 42.9%; Pred. No. 15;
Matches 12; Conservative 8; Mismatches 6; Indels 2; Gaps 1;

Qy      2 GIPCCPVHLKRLIVVVVVVLIIVVING 29
Db      204 GMLASPHYH--LIIIVLVIIILAVVVG 229

RESULT 10
US-10-140-021-340
; Sequence 340, Application US/10140021
; Publication No. US20030138886a1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C167
; CURRENT APPLICATION NUMBER: US/10/140,021
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 340
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-021-340

Query Match          34.5%; Score 59; DB 12; Length 386;
Best Local Similarity 42.9%; Pred. No. 15;
Matches 12; Conservative 8; Mismatches 6; Indels 2; Gaps 1;

Qy      2 GIPCCPVHLKRLIVVVVVVLIIVVING 29
Db      204 GMLASPHYH--LIIIVLVIIILAVVVG 229

RESULT 11
US-10-140-274-340
; Sequence 340, Application US/10140274
; Publication No. US20030143674a1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C163
; CURRENT APPLICATION NUMBER: US/10/140,471
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 340
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-471-340

Query Match          34.5%; Score 59; DB 12; Length 386;
Best Local Similarity 42.9%; Pred. No. 15;
Matches 12; Conservative 8; Mismatches 6; Indels 2; Gaps 1;

Qy      2 GIPCCPVHLKRLIVVVVVVLIIVVING 29
Db      204 GMLASPHYH--LIIIVLVIIILAVVVG 229

RESULT 12
US-10-140-471-340
; Sequence 340, Application US/10140471
; Publication No. US20030138887a1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C163
; CURRENT APPLICATION NUMBER: US/10/140,471
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 340
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-471-340

Query Match          34.5%; Score 59; DB 12; Length 386;
Best Local Similarity 42.9%; Pred. No. 15;
Matches 12; Conservative 8; Mismatches 6; Indels 2; Gaps 1;

Qy      2 GIPCCPVHLKRLIVVVVVVLIIVVING 29
Db      204 GMLASPHYH--LIIIVLVIIILAVVVG 229

RESULT 13
```

```
US-10-140-807-340
; Sequence 340, Application US/10140807
; Publication No. US20030134354A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C174
; CURRENT APPLICATION NUMBER: US/10/140,807
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 340
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-807-340

Query Match          34.5%; Score 59; DB 12; Length 386;
Best Local Similarity 42.9%; Pred. No. 15;
Matches 12; Conservative 8; Mismatches 6; Indels 2; Gaps 1;

OY      2 GIPCCPVHLKRLIVVVVLLIIVVIG 29
Db      204 GMLASPHY--LIIIVLVIIILAVVVG 229

RESULT 14
US-10-140-922-340
; Sequence 340, Application US/10140922
; Publication No. US2003013889A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C179
; CURRENT APPLICATION NUMBER: US/10/140,922
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 340
```

```
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-922-340

Query Match          34.5%; Score 59; DB 12; Length 386;
Best Local Similarity 42.9%; Pred. No. 15;
Matches 12; Conservative 8; Mismatches 6; Indels 2; Gaps 1;

OY      2 GIPCCPVHLKRLIVVVVLLIIVVIG 29
Db      204 GMLASPHY--LIIIVLVIIILAVVVG 229

RESULT 15
US-10-140-924-340
; Sequence 340, Application US/10140924
; Publication No. US20030134355A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C177
; CURRENT APPLICATION NUMBER: US/10/140,924
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 340
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-924-340

Query Match          34.5%; Score 59; DB 12; Length 386;
Best Local Similarity 42.9%; Pred. No. 15;
Matches 12; Conservative 8; Mismatches 6; Indels 2; Gaps 1;

OY      2 GIPCCPVHLKRLIVVVVLLIIVVIG 29
Db      204 GMLASPHY--LIIIVLVIIILAVVVG 229

Search completed: December 3, 2003, 15:56:08
Job time : 21.1905 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using BW model

Run on: December 3, 2003, 15:45:30 ; Search time 10.2381 Seconds  
(without alignments)  
328.763 Million cell updates/sec

Title: US-09-788-308D-3

Sequence: 171  
1 FGIPCCPVHLKRLIVVVVVVIVGALIMGL 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	171	100.0	197	1	LNHUC
2	170	99.4	189	1	LNHUC
3	170	99.4	191	2	LNHUC
4	162	94.7	193	1	A36534
5	162	94.7	194	1	LNHUC
6	150	87.7	35	1	LNHUC
7	148	86.5	34	1	LNHUC
8	140	81.9	190	2	LNHUC
9	136	79.5	35	1	LNHUC
10	136	79.5	35	1	LNHUC
11	60	36.3	193	2	LNHUC
12	58	33.9	165	2	LNHUC
13	58	33.9	165	2	LNHUC
14	57	33.3	166	2	LNHUC
15	56	32.7	88	2	LNHUC
16	56	32.7	105	2	LNHUC
17	56	32.7	534	2	LNHUC
18	55.5	32.5	324	2	LNHUC
19	55	32.2	29	2	LNHUC
20	55	32.2	38	2	LNHUC
21	55	32.2	102	2	LNHUC
22	55	32.2	129	2	LNHUC
23	55	32.2	289	2	LNHUC
24	55	32.2	345	2	LNHUC
25	55	32.2	619	2	LNHUC
26	54.5	31.9	268	2	LNHUC
27	54	31.6	44	2	LNHUC
28	54	31.6	130	2	LNHUC
29	54	31.6	149	2	LNHUC

#### ALIGNMENTS

30	54	31.6	151	2	T36081	probable integral
31	54	31.6	183	2	A64621	hypothetical prote
32	54	31.6	183	2	S52804	virion protein j13
33	54	31.6	185	2	S52903	virion protein j13
34	54	31.6	288	2	A72128	ct345 hypothetical
35	54	31.6	288	2	D81543	hypothetical prote
36	54	31.6	288	2	D86494	CT345 hypothetical
37	54	31.6	289	2	D48213	synthaxin 3 - rat
38	54	31.6	331	2	H83314	NADH dehydrogenase
39	54	31.6	348	2	E82896	hemim perasease U03
40	54	31.6	584	2	I50419	s-glycerin precuro
41	54	31.6	969	2	T33156	hypothetical prote
42	53.5	31.3	1443	2	I50600	neogenin - chicken
43	53	31.0	38	2	PS0118	H-2 class I histoc
44	53	31.0	115	1	DEBGS2	succinate dehydrog
45	53	31.0	115	2	C90722	succinate dehydrog

RESULT 1  
LNHUC  
pulmonary surfactant protein C precursor, long splice form [validated] - human  
N:Alternate names: 3.7 kDa surfactant polypeptide; pulmonary surfactant protein SP5; pul  
N:Contains: pulmonary surfactant protein C precursor, short splice form  
C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1990 #sequence, revision 31-Dec-1990 #text change 08-Dec-2000  
C:Accession: A28801; B28801; I38420; A27338; S02315; S00608; A61249  
R:Glaser, S.W.; Korfhagen, T.R.; Perme, C.M.; Pilot-Matias, T.J.; Klatzer, S.E.; Whitset  
J. Biol. Chem. 263, 10326-10331, 1988  
A:Title: Two Sp-C genes encoding human pulmonary surfactant proteolipid.  
A:Reference number: A28801; PMID:88273133; PMID:2839484  
A:Accession: A28801  
A:Molecule type: DNA  
A:Residues: 1-137, 'T', 139-197 <GLA1>  
A:Cross-references: GB:J03890; NID:G190089; PIDN:AA032022.1; PID:G387029  
A:Experimental source: long splice form  
A:Note: the codon given for residue 138 (ATT) is inconsistent with the authors' translat  
A:Accession: B28801  
A:Molecule type: DNA  
A:Residues: 1-137, 'T', 139-145, 152-197 <GLA2>  
A:Cross-references: GB:J03890; NID:G190089; PIDN:AA032023.1; PID:G387030  
A:Experimental source: short splice form  
A:Note: the codon given for residue 138 (ATT) is inconsistent with the authors' translat  
A:Note: identical proteins are encoded by two genes; transcripts from either gene can be  
R:Hatzi, D.; Deller, G.; Demello, D.E.; Floros, J.  
Exp. Lung Res. 20, 57-72, 1994  
A:Title: Human surfactant protein-C: genetic homogeneity and expression in RDS; comparis  
A:Reference number: I38420; PMID:94237133; PMID:8181452  
A:Accession: I38420  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-14, 'PCQ', 15-44, 'S', 46-64, 'PPQ', 68-137, 'T', 139-185, 'S', 187-197 <HAT>  
A:Cross-references: EMBL:U02948; NID:G498319; PIDN:AA06332.1; PID:G514214  
A:Note: there are probably errors in the presentation of the CDS splice boundaries in th  
R:Glaser, S.W.; Korfhagen, T.R.; Weaver, T.E.; Clark, J.C.; Pilot-Matias, T.; Meuth, J.  
J. Biol. Chem. 263, 9-12, 1988  
A:Title: cDNA, deduced polypeptide structure and chromosomal assignment of human pulmona  
A:Reference number: A27338; PMID:88087156; PMID:333510  
A:Accession: A27338  
A:Molecule type: mRNA  
A:Residues: 1-197 <GLA3>  
A:Cross-references: GB:J03517; NID:G338412; PIDN:AAA36634.1; PID:G338413  
A:Note: part of this sequence, including the amino end of the mature protein, was deter  
R:Warr, R.G.; Hawgood, S.; Buckley, D.I.; Crisp, T.M.; Schilling, J.; Benson, B.J.; Ball  
Proc. Natl. Acad. Sci. U.S.A. 84, 7915-7919, 1987  
A:Title: Low molecular weight human pulmonary surfactant protein (SP5): isolation, chara  
A:Reference number: S02315; PMID:8808508; PMID:3479771  
A:Accession: S02315  
A:Molecule type: mRNA  
A:Residues: 1-197 <WAB>  
A:Cross-references: GB:J03553; NID:G338306; PIDN:AAA36631.1; PID:G338307

A>Note: part of this sequence, including the amino end of the mature protein, was confirmed  
A>Note: 138-Thr and 186-Ser were also found  
R:Johansson, J.; Joernvall, H.; Eklund, A.; Christensen, N.; Robertson, B.; Curstedt, T.  
FEBS Lett. 232, 61-64, 1988  
A:Title: Hydrophobic 3.7 kDa surfactant polypeptide: structural characterization of the  
A:Reference number: S00608; MUID:88211876; PMID:3366248  
A:Accession: S00608  
A:Molecule type: protein  
A:Residues: 24-58 <STU>  
A>Note: 25-Arg was also found  
A>Note: peptides beginning at residues 24, 25, and 26 were detected  
R:Stults, J.T.; Griffin, P.R.; Leskar, D.D.; Naidu, A.; Moffat, B.; Benson, B.J.  
Am. J. Physiol. 261, L118-L125, 1991  
A:Title: Lung surfactant protein SP-C from human, bovine, and canine sources contains pa  
A:Reference number: A61249; MUID:91336436; PMID:1872406  
A:Accession: A61249  
A:Molecule type: protein  
A:Residues: 24-58 <STU>  
A:Note: Identification of palmitoyl cysteines  
C:Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers t  
C:Comment: This protein is synthesized by alveolar type II cells  
C:Comment: The precursor of the hydrophobic proteolipid protein C lacks a typical signal  
C:Genetics:  
A:Gene: GDB:SFTPC; SFTPC2; SP-C  
A:Cross-references: GDB:120373; OMIM:178620  
A:Map position: 8p21-8p21  
A:Introns: 14/3; 67/3; 108/3; 145/3  
A>Note: the first intron occurs before the initiator codon  
C:Superfamily: pulmonary surfactant protein C  
C:Keywords: alternative splicing; gaseous exchange; lipoprotein; lung; pulmonary surfact  
F:1-197/Product: pulmonary surfactant protein C precursor, short splice form #status pr  
F:1-145,152-197/Product: pulmonary surfactant protein C precursor, short splice form #st  
F:1-23/Domain: propeptide #status predicted <PRO>  
F:24-58/Product: pulmonary surfactant protein C #status experimental <MAT>  
F:28/29/Binding site: palmitate (Cys) (covalent) #status experimental

Query Match 100.0%; Score 171; DB 1; Length 197;  
Best Local Similarity 100.0%; Pred. No. 3.9e-12;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGIPCCPVHLKRLIVVVVVVIVVIVGALLMGL 35  
DB 24 FGIPCCPVHLKRLIVVVVVVIVVIVGALLMGL 58

RESULT 2

LNRBC1  
pulmonary surfactant protein C precursor - rabbit  
N:Alternate names: surfactant-associated protein SP-C  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 01-Sep-1995 #sequence revision 25-Apr-1997 #text\_change 18-Feb-2000  
C:Accession: A56766; S14815; A56860; S19946; A56655  
R:Boggarlam, V.; Margana, R.K.  
Am. J. Physiol. 263, L634-L644, 1992  
A:Title: Rabbit surfactant protein C: cDNA cloning and regulation of alternatively splic  
A:Reference number: A56766; MUID:93118799; PMID:1335697  
A:Accession: A56766  
A:Molecule type: mRNA  
A:Residues: 1-189 <BOG>  
A:Cross-references: GB:S51983; NID:g262767; PIDN:AA24761.1; PID:g262768; GB:S51597; NID  
A:Experimental source: lung  
A:Note: sequence extracted from NCBI backbone (NCBIN:121728, NCBI:P.121729)  
A:Note: two types of mRNA were found differing in their 3'-untranslated regions because  
R:Johansson, J.; Persson, P.; Loewenadler, B.; Robertson, B.; Joernvall, H.; Curstedt, T.  
FEBS Lett. 281, 119-122, 1991  
A:Title: Canine hydrophobic surfactant polypeptide SP-C. A lipopeptide with one thioeste  
A:Reference number: S14813; MUID:91200266; PMID:2015882  
A:Accession: S14815  
A:Molecule type: protein  
A:Residues: 24-37, 'XXXXXXXXXXXXXXXXXXXX' <JOH>  
R:Connelly, I.; Possmayer, F.  
Biochim. Biophys. Acta 1127, 199-207, 1992  
A:Title: cDNA sequence and alternative mRNA splicing of surfactant-associated protein C

A:Reference number: A56860; MUID:92353123; PMID:1643107  
A:Accession: A56860  
A:Molecule type: mRNA  
A:Residues: 24-58 <CON>  
A:Cross-references: EMBL:X65078; NID:91720; PIDN:CAA46204.1; PID:g1721  
A:Experimental source: fetal lung  
A:Note: sequence extracted from NCBI backbone (NCBIN:110198, NCBI:P.110234); the complet  
R:Connelly, I.; Possmayer, F.  
submitted to the EMBL Data Library, March 1992  
A:Description: cDNA sequence and alternative splicing of surfactant-associated protein C  
A:Reference number: S19946  
A:Accession: S19946  
A:Molecule type: mRNA  
A:Residues: 3-115,117-161, 'R',163-189 <COC>  
A:Cross-references: EMBL:X65078; NID:91720; PIDN:CAA46204.1; PID:g1721  
R:Durham, P.L.; Nanthakumar, E.U.; Snyder, J.M.  
Exp. Lung Res. 18, 775-793, 1992  
A:Title: Developmental regulation of surfactant-associated proteins in rabbit fetal lung  
A:Reference number: A56655; MUID:93105936; PMID:1468410  
A:Accession: A56655  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 24-115,117-186, 'Y',188-189 <DUR>  
A:Cross-references: GB:S51098; NID:g262066; PIDN:AA24576.1; PID:g262067  
A:Experimental source: fetal lung  
A:Note: sequence extracted from NCBI backbone (NCBIN:121095, NCBI:P.121096); sequence in  
C:Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers t  
C:Comment: This protein is synthesized by alveolar type II cells.  
C:Comment: The precursor of the hydrophobic proteolipid protein C lacks a typical signa  
C:Superfamily: pulmonary surfactant protein C  
C:Keywords: gaseous exchange; lipoprotein; lung; pulmonary surfactant; thiolester bond  
F:24-58/Product: propeptide #status predicted <PRO>  
F:24-58/Product: pulmonary surfactant protein C #status experimental <MAT>  
F:28/29/Binding site: palmitate (Cys) (covalent) #status experimental

Query Match 99.4%; Score 170; DB 1; Length 189;  
Best Local Similarity 97.1%; Pred. No. 4.9e-12;  
Matches 34; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGIPCCPVHLKRLIVVVVVVIVVIVGALLMGL 35  
DB 24 FGIPCCPVHLKRLIVVVVVVIVVIVGALLMGL 58

RESULT 3

G02964  
surfactant protein C - thesus macaque  
C:Species: Macaca mulatta (thesus macaque)  
C:Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 20-Aug-1999  
C:Accession: G02964  
R:An, G.  
submitted to the EMBL Data Library, February 1994  
A:Reference number: G12797  
A:Accession: G02964  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-191 <ANX>  
A:Cross-references: EMBL:U06694; NID:g476267; PIDN:AAA17870.1; PID:g476268  
C:Superfamily: pulmonary surfactant protein C

Query Match 99.4%; Score 170; DB 2; Length 191;  
Best Local Similarity 97.1%; Pred. No. 4.9e-12;  
Matches 34; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGIPCCPVHLKRLIVVVVVVIVVIVGALLMGL 35  
DB 24 FGIPCCPVHLKRLIVVVVVVIVVIVGALLMGL 58

RESULT 4

A36534  
pulmonary surfactant protein C precursor - mouse  
C:Species: Mus musculus (house mouse)



C>Date: 12-Apr-1991 #sequence\_revision 25-Apr-1997 #text\_change 18-Feb-2000  
 C/Accession: A36534  
 R/Glasser, S.W.; Korfthagen, T.R.; Bruno, M.D.; Dey, C.; Whitsett, J.A.  
 J.Biol. Chem. 265, 21986-21991, 1990  
 A/Title: Structure and expression of the pulmonary surfactant protein SP-C gene in the m  
 A/Reference number: A36534; MUID:91072410; PMID:2254341  
 A/Accession: A36534  
 A/Molecule type: DNA  
 A/Residues: 1-193 <GLA>  
 A/Cross-references: GB:M38314; NID:9200559; PIDN:AAA0010.1; PID:9200560  
 C/Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers t  
 C/Comment: This protein is synthesized by alveolar type II cells.  
 C/Comment: The precursor of the hydrophobic proteolipid protein C lacks a typical signal  
 C/Superfamily: pulmonary surfactant protein C  
 C/Keywords: gaseous exchange; lipoprotein; lung; pulmonary surfactant; thiolester bond  
 F:1-23/Domain: propeptide #status predicted <PRO>  
 F:24-58/Product: pulmonary surfactant protein C #status predicted <MAT>  
 F:28,29/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 94.7%; Score 162; DB 1; Length 193;  
 Best Local Similarity 94.3%; Pred. No. 3, 8e-11;  
 Matches 33; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FGIPCCPVHLKRLIVVVVVVLIIVVIVGALLMGL 35  
 Db 24 FRIPCCPVHLKRLIVVVVVVLIIVVIVGALLMGL 58

## RESULT 5

LNRTC  
 pulmonary surfactant protein C precursor - rat  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 18-Feb-2000  
 C/Accession: S03994  
 R/Fisher, J.H.; Shannon, J.M.; Hofmann, T.; Mason, R.J.  
 Biochim. Biophys. Acta 995, 225-230, 1989  
 A/Title: Nucleotide and deduced amino acid sequence of the hydrophobic surfactant protei  
 A/Reference number: S03994; MUID:89207572; PMID:2706272  
 A/Accession: S03994  
 A/Molecule type: mRNA  
 A/Residues: 1-194 <PIS>  
 A/Cross-references: GB:X14221; NID:957286; PIDN:CAA32440.1; PID:957287  
 C/Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers t  
 C/Comment: This protein is synthesized by alveolar type II cells.  
 C/Comment: The precursor of the hydrophobic proteolipid protein C lacks a typical signal  
 C/Superfamily: pulmonary surfactant protein C  
 C/Keywords: gaseous exchange; lipoprotein; lung; pulmonary surfactant; thiolester bond  
 F:1-23/Domain: propeptide #status predicted <PRO>  
 F:24-58/Product: pulmonary surfactant protein C #status predicted <MAT>  
 F:28,29/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 94.7%; Score 162; DB 1; Length 194;  
 Best Local Similarity 94.3%; Pred. No. 3, 8e-11;  
 Matches 33; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FGIPCCPVHLKRLIVVVVVVLIIVVIVGALLMGL 35  
 Db 24 FRIPCCPVHLKRLIVVVVVVLIIVVIVGALLMGL 58

## RESULT 6

LNPGCI  
 pulmonary surfactant protein C [validated] - pig  
 C/Species: Sus scrofa domestica (domestic pig)  
 C/Date: 30-Sep-1989 #sequence\_revision 25-Apr-1997 #text\_change 15-Sep-2000  
 C/Accession: A28640  
 R/Johnson, J.; Curstedt, T.; Robertson, B.; Joernvall, H.  
 Biochemistry 27, 3544-3547, 1988  
 A/Title: Size and structure of the hydrophobic low molecular weight surfactant-associated  
 A/Reference number: A28640; MUID:88309749; PMID:3408709  
 A/Accession: A28640  
 A/Molecule type: protein  
 A/Residues: 1-35 <DOH>

R/Johnson, J.; Szyperki, T.; Curstedt, T.; Muthrich, K.  
 submitted to the Brookhaven Protein Data Bank, September 1994  
 A/Reference number: A52839; EDB:1SPF  
 A/Contents: annotation; conformation by (1)H-NMR, residues 1-35  
 R/Johnson, J.; Szyperki, T.; Curstedt, T.; Muthrich, K.  
 Biochemistry 33, 6015-6023, 1994  
 A/Title: The NMR structure of the pulmonary surfactant-associated polypeptide sp-C in an  
 A/Reference number: A58575; MUID:94235672; PMID:8180229  
 A/Contents: annotation; conformation by (1)H-NMR  
 C/Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers t  
 C/Comment: This protein is synthesized by alveolar type II cells.  
 C/Comment: The precursor of the hydrophobic proteolipid protein C lacks a typical signal  
 C/Superfamily: pulmonary surfactant protein C  
 C/Keywords: gaseous exchange; lipoprotein; lung; pulmonary surfactant; thiolester bond  
 F:5,6/Binding site: palmitate (Cys) (covalent) #status experimental

Query Match 87.7%; Score 150; DB 1; Length 35;  
 Best Local Similarity 90.9%; Pred. No. 2, 8e-10;  
 Matches 30; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IPCCPVHLKRLIVVVVVVLIIVVIVGALLMGL 35  
 Db 3 IPCCPVHLKRLIVVVVVVLIIVVIVGALLMGL 35

## RESULT 7

LNBOCI  
 pulmonary surfactant protein C - bovine  
 N/Alternate names: pulmonary surfactant protein PSP-6  
 C/Species: Bos primigenius taurus (cattle)  
 C/Date: 30-Jun-1989 #sequence\_revision 25-Apr-1997 #text\_change 25-Apr-1997  
 C/Accession: S00609; C61249; S02353; S02318; B29667; S1814  
 R/Johnson, J.; Joernvall, H.; Eklund, A.; Christensen, N.; Robertson, B.; Curstedt, T.  
 FEBS Lett. 232, 61-64, 1988  
 A/Title: Hydrophobic 3.7 kDa surfactant polypeptide: structural characterization of the  
 A/Reference number: S00609; MUID:88211876; PMID:3366248  
 A/Accession: S00609  
 A/Molecule type: protein  
 A/Residues: 1-34 <DOH>  
 R/Stults, J.T.; Griffin, P.R.; Lesikar, D.D.; Naidu, A.; Moffat, B.; Benson, B.J.  
 Am. J. Physiol. 261, E18-E125, 1991  
 A/Title: Lung surfactant protein SP-C from human, bovine, and canine sources contains pa  
 A/Reference number: A61249; MUID:91336436; PMID:1872406  
 A/Accession: C61249  
 A/Molecule type: protein  
 A/Residues: 1-34 <STU>  
 A/Note: Identification of palmitoyl cysteines  
 R/Yu, S.H.; Chung, W.; Olafson, R.W.; Harding, P.G.R.; Possmayer, F.  
 Biochim. Biophys. Acta 921, 437-448, 1987  
 A/Title: Characterization of the small hydrophobic proteins associated with pulmonary su  
 A/Reference number: S02317; MUID:88025156; PMID:3663690  
 A/Accession: S02353  
 A/Molecule type: protein  
 A/Residues: 1-10 <YUI>  
 R/Phelps, D.S.; Smith, L.M.; Taeusch, H.W.  
 Am. Rev. Respir. Dis. 135, 1112-1117, 1987  
 A/Title: Characterization and partial amino acid sequence of a low molecular weight surf.  
 A/Reference number: S02318; MUID:87211387; PMID:3579010  
 A/Accession: S02318  
 A/Molecule type: protein  
 A/Residues: 1-3, 'X', 6-10, 'L', 12-15, 'X', 17-21, 'I', <PHE>  
 R/Olafson, R.W.; Rink, U.; Kjelland, S.; Yu, S.H.; Chung, J.; Harding, P.G.R.; Possmayer  
 Biochim. Biophys. Res. Commun. 148, 1406-1411, 1987  
 A/Title: Protein sequence analysis studies on the low molecular weight hydrophobic prote  
 A/Reference number: A50137; MUID:88077030; PMID:3689402  
 A/Accession: B29667  
 A/Molecule type: protein  
 A/Residues: 1-20, 'V', 22-25, 'V', 27, 'IGAMLA', <OLA>  
 R/Johnson, J.; Persson, P.; Loewenadler, B.; Robertson, B.; Joernvall, H.; Curstedt, T.  
 FEBS Lett. 281, 119-122, 1991  
 A/Title: Canine hydrophobic surfactant polypeptide SP-C. A lipopeptide with one thioeste  
 A/Reference number: S14813; MUID:91200266; PMID:2015882  
 A/Contents: annotation

C/Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers C/Comment: This protein is synthesized by alveolar type II cells.  
C/Comment: The precursor of the hydrophobic proteolipid protein C lacks a typical signal C/Superfamily: pulmonary surfactant protein C  
C/Keywords: gaseous exchange; lipoprotein; lung; pulmonary surfactant; cholesterol bond F/4.5/Binding site: palmitate (Cys) (covalent) #status experimental

Query Match 86.5%; Score 148; DB 1; Length 34;  
Best Local Similarity 90.9%; Pred. No. 4.5e-10;  
Matches 30; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IPCCPVHLKRLIVVVVVVIVGALLMGL 35  
Db 2 IPCCPVHLKRLIVVVVVVIVGALLMGL 34

RESULT 8  
S31490  
surfactant protein SP-C - American mink

C/Species: Mustela vison (American mink)  
C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Aug-1999

C/Accession: S31490

R/Christensen, J.; Belousov, J.; Storgaard, T.; Aasted, B.; Alexandersen, S.  
submitted to the EMBL Data Library, December 1992

A/Description: Nucleotide and deduced amino acid sequence of the hydrophobic surfactant A/Reference number: S31490

A/Accession: S31490

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-190 <CHR>

A/Cross-references: EMBL:Z19516; NID:G1188; PIDN:CAA79577.1; PID:G1189

C/Superfamily: pulmonary surfactant protein C

Query Match 81.9%; Score 140; DB 2; Length 190;  
Best Local Similarity 80.0%; Pred. No. 1e-08;  
Matches 28; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 FGIPCCPVHLKRLIVVVVVVIVGALLMGL 35  
Db 24 FGIPCCPVHLKRLIVVVVVVIVGALLMGL 58

RESULT 9  
LNDGCI

pulmonary surfactant protein C - dog

C/Species: Canis lupus familiaris (dog)

C/Date: 12-May-1994 #sequence\_revision 25-Apr-1997 #text\_change 25-Apr-1997

C/Accession: B61249; S14813

R/Stulcs, J.T.; Griffin, P.R.; Lesikar, D.D.; Naidu, A.; Moffatt, B.; Benson, B.J.  
Am. J. Physiol. 261, L118-L125, 1991

A/Title: lung surfactant protein SP-C from human, bovine, and canine sources contains pe A/Reference number: A61249; MUID:91336436; PMID:1872406

A/Accession: B61249

A/Molecule type: protein

A/Residues: 1-35 <STU>

R/Johansson, J.; Persson, P.; Loewenadler, B.; Robertson, B.; Joernvall, H.; Curstedt, T  
FEBS Lett. 281, 119-122, 1991

A/Title: Canine hydrophobic surfactant polypeptide SP-C. A lipopeptide with one thioeste A/Reference number: S14813; MUID:91200266; PMID:2015882

A/Accession: S14813

A/Molecule type: protein

A/Residues: 2-35 <JOH>

C/Comment: Pulmonary surfactant is a complex of phospholipids and proteins that lowers C/Comment: This protein is synthesized by alveolar type II cells.

C/Comment: The precursor of the hydrophobic proteolipid protein C lacks a typical signal C/Superfamily: pulmonary surfactant protein C

C/Keywords: gaseous exchange; lipoprotein; lung; pulmonary surfactant; cholesterol bond F/5/Binding site: palmitate (Cys) (covalent) #status experimental

Query Match 79.5%; Score 136; DB 1; Length 35;  
Best Local Similarity 82.4%; Pred. No. 9.9e-09;  
Matches 28; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GIPCCPVHLKRLIVVVVVVIVGALLMGL 35  
Db 2 GIPCCPVHLKRLIVVVVVVIVGALLMGL 35

RESULT 10

F72311  
lema protein - Thermotoga maritima (strain MSB8)

C/Species: Thermotoga maritima

C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C/Accession: F72311

R/Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.

Nature 399, 323-329, 1999

A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq A/Reference number: A72200; MUID:99287316; PMID:10360571

A/Accession: F72311

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-193 <ARN>

A/Cross-references: GB:AE001759; GB:AE000512; NID:94981495; PIDN:AAD36040.1; PID:9498145

A/Experimental source: strain MSB8

C/Genetics:

A/Gene: TM0961

Query Match 36.3%; Score 62; DB 2; Length 193;  
Best Local Similarity 50.0%; Pred. No. 4.5;  
Matches 11; Conservative 9; Mismatches 2; Indels 0; Gaps 0;

Qy 8 VHLKRLIVVVVVVIVG 29  
Db 5 IRVKGRLIVLVIVLVIG 26

RESULT 11

E72752

hypothetical protein APE0010 - Aeropyrum pernix (strain KI)

C/Species: Aeropyrum pernix

C/Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999

C/Accession: E72752

R/Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hakiyama, Y.; Jin-no, K.; Takat  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; I  
DNA Res. 6, 83-101, 1999

A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr A/Reference number: A72450; MUID:99310339; PMID:10382966

A/Accession: E72752

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-538 <KAW>

A/Cross-references: DDBJ:AF000058; NID:95103388; PIDN:BAW78919.1; PID:dl042695; PID:9510

A/Experimental source: strain KI

C/Genetics:

A/Gene: APE0010

Query Match 35.1%; Score 60; DB 2; Length 538;  
Best Local Similarity 40.6%; Pred. No. 14;  
Matches 13; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

Qy 3 IPCCPVHLKRLIVVVVVVIVGALLMGL 34  
Db 121 INCCPREAYIALVFPVAGLVVVMGLMSG 152

RESULT 12

B82670

general secretory pathway protein G precursor XF1519 (imported) - Xylella fastidiosa (st

C/Species: Xylella fastidiosa

C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 17-Nov-2000

C/Accession: B82670

R/anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen Nature 406, 151-157, 2000

A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; NCID:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: B82670  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-165 <SIM>  
A:Cross-references: GB:AE003982; GB:AE003849; NID:G9106554; PIDN:AAF84328.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Brienes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, H  
as-Neto, E.; Docena, C.; El-Dorty, H.; Facincani, A.F.; Ferreira, A.J.S.  
submitted to Genbank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Klitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laizy  
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miranda, E.C.; Miyaki, C.Y.  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeiri, D.A  
Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv  
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XFI519  
C:Superfamily: secretion protein xcpT  
Query Match 33.9%; Score 58; DB 2; Length 165;  
Best Local Similarity 34.2%; Pred. No. 11;  
Matches 13; Conservative 11; Mismatches 8; Indels 6; Gaps 2;  
OY 3 IPCCPVHLKLR---LIIVVVVVLI--VVIVGALLMG 34  
DB 28 ITCSPQMRQAGMSLEIIVIVLIGGVIAFVGSRLVG 65  
RESULT 13  
B83778  
Ferrichrome ABC transporter (permease) BHI026 [imported] - Bacillus halodurans (strain C  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_01-Dec-2000 #text\_change 17-May-2002  
C:Accession: B83778  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hir  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; NCID:20512582; PMID:11058132  
A:Accession: B83778  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-315 <STO>  
A:Cross-references: GB:AP001510; GB:BA000004; NID:G10173440; PIDN:BA804745.1; GSPDB:GN00  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BHI026  
C:Superfamily: ferrichrome ABC transporter  
Query Match 33.9%; Score 58; DB 2; Length 315;  
Best Local Similarity 45.7%; Pred. No. 17;  
Matches 16; Conservative 6; Mismatches 11; Indels 2; Gaps 2;  
OY 2 GIPCCPVHLKRLIIIVVVVLIIVVIVGAL-LMGL 35  
DB 212 GIPYDSV-IKRLIIIVAIISITATLVGPITFLGL 245  
RESULT 14  
B82873  
hypothetical protein UT580 [imported] - Ureaplasma urealyticum  
C:Species: Ureaplasma urealyticum  
C:Date: 18-Aug-2000 #sequence\_20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: H82873  
R:Glas, J.I.; Lefkowitz, E.J.; Glas, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.  
submitted to GenBank, February 2000  
A:Description: The complete sequence of Ureaplasma urealyticum. Alternate views of a mi

A:Reference number: A82870  
A:Accession: H82873  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-166 <GLA>  
A:Cross-references: GB:AE002156; GB:AF222894; NID:G6899580; PIDN:AAF30994.1; GSPDB:GN001  
A:Experimental source: serovar 3; biovar 1  
C:Genetics:  
A:Gene: UT580  
A:Genetic code: SGC3  
Query Match 33.3%; Score 57; DB 2; Length 166;  
Best Local Similarity 32.0%; Pred. No. 14;  
Matches 8; Conservative 12; Mismatches 5; Indels 0; Gaps 0;  
OY 11 KRLLIVVVVLIIVVIVGALLMG 35  
DB 132 KRSVILIIIVVILIAAIFFGI 156  
RESULT 15  
T47139  
hypothetical protein DKFZp761P2414.1 - human  
C:Species: Homo sapiens (man)  
C:Date: 20-Apr-2000 #sequence\_20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T47139  
R:Angorge, W.; Winkler, U.; Mewes, H.W.; Weil, B.; Wiemann, S.  
submitted to the Protein Sequence Database, March 2000  
A:Reference number: Z24375  
A:Accession: T47139  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-88 <AAA>  
A:Cross-references: EMBL:AL161976  
A:Experimental source: adult amygdala; clone DKFZp761P2414  
C:Genetics:  
A:Note: DKFZp761P2414.1  
Query Match 32.7%; Score 56; DB 2; Length 88;  
Best Local Similarity 45.0%; Pred. No. 14;  
Matches 9; Conservative 7; Mismatches 4; Indels 0; Gaps 0;  
OY 15 IYVVVVVLIIVVIVGALLMG 34  
DB 12 VTVVIVILIAALGALLIG 31

Search completed: December 3, 2003, 15:53:28  
Job time : 11.2381 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 15:43:05 ; Search time 6.19048 Seconds  
(without alignments)

265.882 Million cell updates/sec

Title: US-09-788-308d-3  
Perfect score: 171  
Sequence: 1 FGIPCCPVHLKRLIVVVVVVIVIVIGALLMGL 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171	100.0	197	1	PSPC_HUMAN
2	170	99.4	188	1	PSPC_RABIT
3	170	99.4	191	1	PSPC_MACMU
4	162	94.7	193	1	PSPC_MOUSE
5	162	94.7	194	1	PSPC_RAT
6	150	87.7	35	1	PSPC_PIG
7	148	86.5	34	1	PSPC_BOVIN
8	140	81.9	190	1	PSPC_MUSVI
9	136	79.5	34	1	PSPC_CANFA
10	59	34.5	386	1	T10D_HUMAN
11	56	32.7	88	1	SNN_HUMAN
12	56	32.7	88	1	SNN_MOUSE
13	56	32.7	105	1	YIF4_YEAST
14	55	32.2	289	1	STX3_MOUSE
15	54	31.6	289	1	STX3_RAT
16	53.5	31.3	1377	1	NEO1_RAT
17	53.5	31.3	1443	1	NEO1_CHICK
18	53.5	31.3	1461	1	NEO1_HUMAN
19	53	31.0	115	1	DHSD_ECOLI
20	53	31.0	321	1	VP05_YACCC
21	53	31.0	322	1	VP05_VACCV
22	53	31.0	322	1	BMRB_MOUSE
23	52.5	30.7	1493	1	NEO1_MOUSE
24	52	30.4	289	1	STX3_HUMAN
25	52	30.4	474	1	CTF5_HUMAN
26	52	30.4	636	1	CTR4_HUMAN
27	52	30.4	1310	1	AEB3_HUMAN
28	51.5	30.1	468	1	MLEN_BACSU
29	51.5	30.1	565	1	VGIF_SENDF
30	51.5	30.1	565	1	VGIF_SENDF
31	51	29.8	220	1	SYBL_HUMAN
32	51	29.8	246	1	ENV_RSVSR
33	51	29.8	287	1	THEF_HUMAN

34	51	29.8	306	1	C181_HUMAN
35	51	29.8	357	1	HA15_MOUSE
36	51	29.8	565	1	VGIF_SENDF
37	51	29.8	565	1	VGIF_SENDF
38	51	29.8	586	1	PMEL_ARATH
39	51	29.8	740	1	PECI_PIG
40	50.5	29.5	256	1	YDH1_HSVSC
41	50.5	29.5	365	1	GRAB_BACSU
42	50.5	29.5	754	1	AD07_HUMAN
43	50.5	29.5	4660	1	LRP2_RAT
44	50	29.2	174	1	ENV_AVISU
45	50	29.2	191	1	TXIA_SYNP7

## ALIGNMENTS

RESULT 1  
PSPC\_HUMAN  
ID PSPC\_HUMAN STANDARD: PRT: 197 AA.  
AC P1686; P1687; Q12793;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Pulmonary surfactant-associated protein C precursor (SP-C) (SP5)  
DE (Pulmonary surfactant-associated proteolipid SPL(Val)).  
GN SPFP2 OR SPFP2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP MEDLINE=88087156; PubMed=3335510;  
RX SEQUENCE FROM N.A., AND SEQUENCE OF 26-42.  
RA Glasser S.W., Korfagen T.R., Weaver T.E., Clark J.C.,  
RA Pilot-Matias T., Meuth J., Fox J.L., Whitsett J.A.;  
RT "cDNA, deduced polypeptide structure and chromosomal assignment of  
RT human pulmonary surfactant proteolipid, SPL(Val).";  
RL J. Biol. Chem. 263:9-12(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88273133; PubMed=2839484;  
RA Glasser S.W., Korfagen T.R., Perme C.M., Pilot-Matias T.J.,  
RA Kiester S.B., Whitsett J.A.;  
RT "Two SP-C genes encoding human pulmonary surfactant proteolipid.";  
RL J. Biol. Chem. 263:10326-10331(1988).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88068508; PubMed=3479771;  
RA Warr R.G., Hawgood S., Buckley D.I., Crisp T.M., Schilling J.,  
RA Benson B.J., Ballard P.L., Clements J.A., White R.T.;  
RT "Low molecular weight human pulmonary surfactant protein (SP5):  
RT isolation, characterization, and cDNA and amino acid sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:7915-7919(1987).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94237133; PubMed=8181452;  
RA Hatzis D., Dieter G., demello D.E., Floros J.;  
RT "Human surfactant protein-C: genetic homogeneity and expression in  
RT RDS; comparison with other species.";  
RL Exp. Lung Res. 20:57-72(1994).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX TISSUE=LUNG;  
RX MEDLINE=22386257; PubMed=12477932;  
RA Strauberg R.L., Feinberg E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Dlatkovich L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stepieton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carinici P., Prange C.,  
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,



```
CC ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-
CC LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- MISCELLANEOUS: PULMONARY SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

EMBL; X65078; CAA46204.1; ALT\_INIT.

EMBL; S51983; AAB24761.1; -.

EMBL; AF037445; AAC18032.1; -.

EMBL; S51597; AAB24762.1; -.

EMBL; S51098; AAB24576.2; -.

PIR; A56766; LNRBCL.

InterPro; IPR001729; Pulm\_surfact\_AP.

Pfam; PF04089; BRICHOS; 1.

ProDom; PD009591; Pulm\_surfact\_AP; 1.

SMART; SM00019; SP\_P; 1.

PROSITE; PS00341; SURFACT\_PALMTOYL; 1.

Surface film; Gaseous exchange; Lipoprotein; Palmitate.

PROPEP

CHAIN 1 23

PROPEP 59 188

LIPID 28 28

LIPID 29 29

CONFLICT 115 115

CONFLICT 153 153

CONFLICT 159 159

CONFLICT 161 161

CONFLICT 186 186

SEQUENCE 188 AA; 19836 MW; F622EEA933786678 CRC64;

Query Match 99.4%; Score 170; DB 1; Length 188;  
Best Local Similarity 97.1%; Pred. No. 3e-12;  
Matches 34; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 FGIPCCPHLKRLIVVVVVVIVVIVGALLMGL 35  
24 FGIPCCPHLKRLIVVVVVVIVVIVGALLMGL 58

RESULT 3

PSPC\_MOUSE STANDARD; PRT; 191 AA.

AC P55152;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Pulmonary surfactant-associated protein C precursor (SP-C) (Pulmonary surfactant-associated proteolipid SPL(Val)).

GN SFTPC OR SFTP2.

OS Maccaca mulatta (rhesus macaque).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Macaca.

NCBI\_TaxID=9544;

NCBI\_TaxID=9544;

SEQUENCE FROM N.A.

RC TISSUE=Lung;

RA An G., Luo G., Zhao Y., Plopper C., Wu R.;

RU Submitted (May-1994) to the EMBL/Genbank/DBJ databases.

CC -1- FUNCTION: PULMONARY SURFACTANT ASSOCIATED PROTEINS PROMOTE ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES.

```
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- MISCELLANEOUS: PULMONARY SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

EMBL; U06694; AAA17870.1; -.

PIR; G02964; G02964.

InterPro; IPR001729; Pulm\_surfact\_AP.

Pfam; PF04089; BRICHOS; 1.

ProDom; PD009591; Pulm\_surfact\_AP; 1.

SMART; SM00019; SP\_P; 1.

PROSITE; PS00341; SURFACT\_PALMTOYL; 1.

Surface film; Gaseous exchange; Lipoprotein; Palmitate.

PROPEP

CHAIN 1 23

PROPEP 59 191

LIPID 28 28

LIPID 29 29

CONFLICT 115 115

CONFLICT 153 153

CONFLICT 159 159

CONFLICT 161 161

CONFLICT 186 186

SEQUENCE 191 AA; 20584 MW; 4A653FB3EB94C02D CRC64;

Query Match 99.4%; Score 170; DB 1; Length 191;  
Best Local Similarity 97.1%; Pred. No. 3e-12;  
Matches 34; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 FGIPCCPHLKRLIVVVVVVIVVIVGALLMGL 35  
24 FGIPCCPHLKRLIVVVVVVIVVIVGALLMGL 58

RESULT 4

PSPC\_MOUSE STANDARD; PRT; 193 AA.

AC P21841;

DT 01-MAY-1991 (Rel. 18, Created)

DT 01-MAY-1991 (Rel. 18, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Pulmonary surfactant-associated protein C precursor (SP-C) (SPS) (Pulmonary surfactant-associated proteolipid SPL(Val)).

GN SFTPC OR SFTP2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_TaxID=10090;

NCBI\_TaxID=10090;

SEQUENCE FROM N.A.

RC MEDLINE=91072410; PubMed=2254341;

RA Glaeser S.W., Korenegen T.R., Bruno M.D., Dey C., Whitsett J.A.;

RT "Structure and expression of the pulmonary surfactant protein SP-C gene in the mouse.";

RU J. Biol. Chem. 265:21986-21991(1990).

CC -1- FUNCTION: PULMONARY SURFACTANT ASSOCIATED PROTEINS PROMOTE ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES.

CC -1- SUBCELLULAR LOCATION: Extracellular.

CC -1- MISCELLANEOUS: PULMONARY SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS, CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL HYDROPHOBIC PROTEINS (SP-B AND SP-C).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL; M88314; AAA40010.1; -  
DR PIR; A36534; A36534.  
DR MGD; MG1:109517; Sftpc.  
DR InterPro: IPR001729; Pulm\_surfact\_AP.  
DR Pfam; PF04089; BRICHOS; 1. Pulm\_surfact\_AP.  
DR ProDom; PD009591; Pulm\_surfact\_AP; 1.  
DR SMART; SM000019; SP\_P; 1.  
DR POSITIVE; PS00341; SURFACT PALMYTOYL; 1.  
KM Surface film; Gaseous exchange; Lipoprotein; Palmitate.  
FT PROPEP 1 23  
FT CHAIN 24 58  
FT PROPEP 59 193  
FT LIPID 28 28  
FT LIPID 29 29  
SQ SEQUENCE 193 AA; 21054 MW; 10CCD91889AC8301 CRC64;

Query Match 94.7%; Score 162; DB 1; Length 193;  
Best Local Similarity 94.3%; Pred. No. 2.3e-11;  
Matches 33; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1 FGIPCCPVHLKRLIIVVVVIVVIVGALLMGL 35  
24 FRIPCCPVHLKRLIIVVVVIVVIVGALLMGL 58

RESULT 5  
PSPC\_RAT ID STANDARD; PRT: 194 AA.  
AC P1685;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Pulmonary surfactant-associated protein C precursor (SP-C) (Pulmonary  
DE surfactant-associated proteolipid SPL(Val)).  
GN SFTPC OR SFTP2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus;  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=Sprague-Dawley; TISSUE=Lung;  
RX MEDLINE=89207572; PubMed=2706272;  
RA Fisher J.H., Shannon J.M., Hatmann T., Mason R.J.;  
RT "Nucleotide and deduced amino acid sequence of the hydrophobic  
RT surfactant protein SP-C from rat: expression in alveolar type II  
RT cells and homology with SP-C from other species.";  
RL Biochim. Biophys. Acta 995:225-230(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Rishi A.K., Gulamhussein A.I., Albanese S., Williams M.C.,  
RA Brody J.S.;  
RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: PULMONARY SURFACTANT ASSOCIATED PROTEINS PROMOTE  
CC ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-  
CC LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES.  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%  
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,  
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL  
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
DR EMBL; X14221; CAA32440.1; -  
DR PIR; S03994; LNRTC.  
DR InterPro: IPR001729; Pulm\_surfact\_AP.  
DR Pfam; PF04089; BRICHOS; 1. Pulm\_surfact\_AP.  
DR ProDom; PD009591; Pulm\_surfact\_AP; 1.  
DR SMART; SM000019; SP\_P; 1.  
DR POSITIVE; PS00341; SURFACT PALMYTOYL; 1.  
KM Surface film; Gaseous exchange; Lipoprotein; Palmitate.  
FT PROPEP 1 23  
FT CHAIN 24 58  
FT PROPEP 59 194  
FT LIPID 28 28  
FT LIPID 29 29  
FT CONFLICT 188 188  
SQ SEQUENCE 194 AA; 21042 MW; 51A75A18E4C95867 CRC64;

Query Match 94.7%; Score 162; DB 1; Length 194;  
Best Local Similarity 94.3%; Pred. No. 2.3e-11;  
Matches 33; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

1 FGIPCCPVHLKRLIIVVVVIVVIVGALLMGL 35  
24 FRIPCCPVHLKRLIIVVVVIVVIVGALLMGL 58

RESULT 6  
PSPC\_PIG ID STANDARD; PRT: 35 AA.  
AC P15785;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Pulmonary surfactant-associated proteolipid C (SP-C) (Pulmonary  
DE surfactant-associated proteolipid SPL(Val)).  
GN SFTPC OR SFTP2.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus;  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=88309749; PubMed=3408709;  
RA Johanson J., Curreted T., Robertson B., Joernvall H.;  
RT "Size and structure of the hydrophobic low molecular weight  
RT surfactant-associated polypeptide.";  
RL Biochemistry 27:3544-3547(1988).  
RN [2]  
RP PALMITOYLATION.  
RX MEDLINE=9022154; PubMed=2326260;  
RA Curreted T., Johanson J., Persson P., Eklund A., Robertson B.,  
RA Loewenadler B., Joernvall H.;  
RT "Hydrophobic surfactant-associated polypeptides: SP-C is a  
RT lipopeptide with two palmitoylated cysteine residues, whereas SP-B  
RT lacks covalently linked fatty acyl groups.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:2985-2989(1990).  
RN [3]  
RP STRUCTURE BY NMR.  
RX MEDLINE=94235672; PubMed=8180229;  
RA Uchanson J., Szyperaki T., Curreted T., Wuehrich K.;  
RT "The NMR structure of the pulmonary surfactant-associated polypeptide  
RT SP-C in an apolar solvent contains a valyl-rich alpha-helix.";  
RL Biochemistry 33:6015-6023(1994).  
CC -1- FUNCTION: PULMONARY SURFACTANT ASSOCIATED PROTEINS PROMOTE  
CC ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-  
CC LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES.  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%  
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,

CC	CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC	HYDROPHOBIC PROTEINS (SP-B AND SP-C).
DR	PIR: A28640; INPGL1.
DR	PDB: 1SPF; 20-DEC-94.
DR	InterPro: IPR001729; Pulm surfact AP.
DR	ProSITE: PS00341; SURFACT_PALMYTOL; 1.
KW	Surface film; Gaseous exchange; Lipoprotein; Palmitate; 3D-structure.
FT	LIPID 5 5 PALMITATE.
FT	LIPID 5 6 PALMITATE.
SO	SEQUENCE 35 AA; 3710 MW; C6A713AF926FOFE CRC64;
Oy	Query Match 87.7%; Score 150; DB 1; Length 35;
Db	Best Local Similarity 90.9%; Pred. No. 1.4e-10;
	Matches 30; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
	3 IPCCPVHLKRLIVVVVVLIVVVGALLMGL 35
	3 IPCCVNLKRLVVVVVVVLLVVGALLMGL 35
RESULT 7	
PSPC BOVIN	STANDARD; PRT; 34 AA.
ID	P15783; P15784; Created)
AC	01-APR-1990 (Rel. 14,
DT	01-APR-1990 (Rel. 14, Last sequence update)
DT	15-JUL-1999 (Rel. 38, Last annotation update)
DE	Pulmonary surfactant-associated protein C (SP-C) (Pulmonary
GN	surfactant-associated proteolipid SPL(Val)).
OS	SPPRC OR SPPRC2.
OC	Bos taurus (Bovine).
OC	Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OX	Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;	[1]
RN	SEQUENCE.
RP	MEDLINE=86211876; PubMed=3366248;
RX	Johansson J., Joernvall H., Eklund A., Christensen N., Robertsson B.,
RA	Curstedt T.;
RT	"Hydrophobic 3.7 kDa surfactant polypeptide: structural
RT	characterization of the human and bovine forms.";
RL	FEMS Lett. 232:61-64(1988).
RN	[2]
RP	SEQUENCE.
RX	MEDLINE=86077030; PubMed=3689402;
RA	Olaesson R.W., Rink U., Kiehlend S., Yu S.-H., Chung J.,
RA	Harding P.G.R., Possmayer F.;
RT	"Protein sequence analysis studies on the low molecular weight
RT	hydrophobic proteins associated with bovine pulmonary surfactant."
RL	Biochim. Biophys. Res. Commun. 148:1406-1411(1987).
RN	[3]
RP	SEQUENCE OF 1-10.
RX	MEDLINE=86025156; PubMed=3636390;
RA	Yu S.-H., Chung W., Olaesson R.W., Harding P.G.R., Possmayer F.;
RT	"Characterization of the small hydrophobic proteins associated with
RT	pulmonary surfactant."
RL	Biochim. Biophys. Acta 921:437-448(1987).
RN	[4]
RP	PALMITOYLATION.
RX	MEDLINE=91336436; PubMed=1872406;
RA	Stults J.T., Griffin P.R., Lesker D.D., Naidu A., Moffat B.,
RA	Benson B.J.;
RT	"Lung surfactant protein SP-C from human, bovine, and canine sources
RT	contains palmitoyl cysteine thioester linkages.";
RL	Am. J. Physiol. 261:L118-L125(1991).
CC	-1- FUNCTION: PULMONARY SURFACTANT ASSOCIATED PROTEINS PROMOTE
CC	ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-
CC	-LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES.
CC	-1- SUBCELLULAR LOCATION: Extracellular
CC	-1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC	PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC	CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL

```
CC      HYDROPHOBIC PROTEINS (SP-B AND SP-C).
DR      PIR; S00609; LNBCL.
DR      InterPro; IPR001729; Pulm_surface_AP.
DR      ProSITE; PS00341; SURFACT_PALMYTOYL; 1.
KW      Surface film; Gaseous exchange; Lipoprotein; Palmitate.
FT      LIPID          4         4
FT      CONFLICT       5         5
FT      CONFLICT       21        21
FT      CONFLICT       26        26
FT      CONFLICT       28        28
SQ      SEQUENCE      34 AA; 3582 MW; C8B031E9C0F820FE CRC64;
QY      Query Match      86.5%; Score 148; DB 1; Length 34;
      Best Local Similarity 90.9%; Pred. No. 2; 3-10;
      Matches 30; Conservative 3; Mismatches 0; Indels 0; Gaps 0.
Db      3 IPCCPVHLKRLIIVVVVVLLIVVIVGALLMGL 35
      2 IPCCPVNIKRLIIVVVVLLIIVVIGALLMGL 34

RESULT 8
BPSC_MUSVI STANDARD; PRT; 190 AA.
AC P35245;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pulmonary surfactant-associated protein C precursor (SP-C).
OS SpTpc OR SpT2P.
OC Musceta vison (American mink).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;
OC Mustela.
OX NCBI_TaxID=9667;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Christensen J., Belusov J., Storgaard T., Aasted B., Alexandersen S.;
RV Submitted (DEC-1992) to the EMBL/GenBank/DDBJ databases.
CC -! FUNCTION: PULMONARY SURFACTANT ASSOCIATED PROTEIN PROMOTE
CC ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-
CC LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES.
CC -! SUBCELLULAR LOCATION: Extracellular.
CC -! MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC DR EMBL; Z19516; CAAT9577.1; -.
CC DR PIR; S31490; S31490.
CC DR InterPro; IPR001729; Pulm_surface_AP.
CC DR Pfam; PF04089; BRICHOS_1.
CC DR ProDom; PD009591; Pulm_surfact_AB; 1.
CC DR SMART; SM00019; SF_P; 1.
CC DR PROSITE; PS00341; SURFACT_PALMYTOYL; FALSE_NEG.
KW Surface film; Gaseous exchange; Lipoprotein; Palmitate.
FT FT PROPEP 1 23
FT CHAIN 24 58
FT PROPEP 59 190
FT LIPID 28 PALMITATE (BY SIMILARITY).
SO SEQUENCE 190 AA; 20431 MW; FFAC62747A11E27 CRC64;
Query Match 81.9%; Score 140; DB 1; Length 190;
```



Best Local Similarity 80.0%; Pred. No. 5.9e-09;  
Matches 28; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 FGIPCCVHLKRLIVVVVVIVVIVGALLMGL 35  
Db 24 FGIPCCVHLKRLIVVVVVIVVIVGALLMGL 58

RESULT 9  
PSPC\_CANFA STANDARD; PRT; 34 AA.  
ID PSPC\_CANFA  
AC P22397;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Pulmonary surfactant-associated protein C (SP-C) (Pulmonary  
DE surfactant-associated proteolipid SPL (Val)).  
GN SFTPC OR SFTP2.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN (1)  
RP SEQUENCE.  
RX MEDLINE=91200266; PubMed=2015882;  
RA Johanson J., Persson P., Loewenadler B., Robertson B., Joernvall H.,  
RA Curstedt T.;  
RT "Canine hydrophobic surfactant polypeptide SP-C. A lipopeptide with  
RT one cholesterol-linked palmitoyl group.";  
RL FEBS Lett. 281:119-122(1991).  
CC -1- FUNCTION: PULMONARY SURFACTANT ASSOCIATED PROTEINS PROMOTE  
CC ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-  
CC LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES.  
CC -1- SUBCELLULAR LOCATION: Extracellular.  
CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%  
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,  
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL  
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).  
DR InterPro: IPR001729; Ptm surfact AP.  
DR PIR: B61249; LINDGCL.  
DR PROSITE: PS00341; SURFACT\_PALMITOYL\_FALSE\_NEG.  
KW Surface film; Gaseous exchange; Lipoprotein; Palmitate.  
FT Lipid 4 4 PALMITATE.  
SQ SEQUENCE 34 AA; 3545 MW; 246AD8847DBA70B3 CRC64;

Query Match 79.5%; Score 136; DB 1; Length 34;  
Best Local Similarity 82.4%; Pred. No. 4.8e-09;  
Matches 28; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 GIPCCVHLKRLIVVVVVIVVIVGALLMGL 35  
Db 1 GIPCCVHLKRLIVVVVVIVVIVGALLMGL 34

RESULT 10  
ID T10D\_HUMAN STANDARD; PRT; 386 AA.  
AC Q9UBN6; Q9Y604;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Tumor necrosis factor receptor superfamily member 10D precursor (Decoy  
DE receptor 2) (DCR2) (TNF-related apoptosis-inducing ligand receptor 4)  
DE (TRAIL receptor-4) (TRAIL-R4) (TRAIL receptor with a truncated death  
DE domain).  
GN TNFRSF10D OR DCR2 OR TRAILR4 OR TRUND.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A., SEQUENCE OF N-TERMINUS, AND VARIANT.  
RN RP  
RC TISSUE=Fetal lung;

RA MEDLINE=98044290; PubMed=9382840;  
RA Mersereau S.A., Sheridan J.P., Pitti R.M., Huang A., Skubatch M.,  
RA Baldwin D., Yuan J., Gurney A., Goddard A.D., Godowski P.,  
RA Ashkenazi A.;  
RT "A novel receptor for Apo2L/TRAIL contains a truncated death domain.";  
RL Curr. Biol. 7:1003-1006(1997).  
RN (2)  
RP SEQUENCE FROM N.A., CHARACTERIZATION AND VARIANTS SER-35 AND LEU-310.  
RC TISSUE=Forebrain fibroblast, and Peripheral blood lymphocytes;  
RX MEDLINE=98090091; PubMed=9430226;  
RA Degli-Esposti M.A., Dougall W.C., Smolak P.J., Waugh J.Y., Smith C.A.,  
RA Goodwin R.G.;  
RT "The novel receptor TRAIL-R4 induces NF-kappaB and protects against  
RT TRAIL-mediated apoptosis, yet retains an incomplete death domain.";  
RL Immunity 7:813-820(1997).  
RN (3)  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RX MEDLINE=98196860; PubMed=9537512;  
RA Pan G., Ni J., Yu G.-L., Wei Y.-F., Dixit V.M.;  
RT "TRUND, a new member of the TRAIL receptor family that antagonizes  
RT TRAIL signalling.";  
RL FEBS Lett. 424:41-45(1998).  
CC -1- FUNCTION: Receptor for the cytotoxic ligand TRAIL. Contains a  
CC truncated death domain and hence is not capable of inducing  
CC apoptosis but protects against TRAIL-mediated apoptosis. Reports  
CC are contradictory with regards to its ability to induce the NF-  
CC kappaB pathway (According to Ref.1 it cannot but according to  
CC Ref.2 it can induce the NF-kappaB pathway).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- TISSUE SPECIFICITY: Widely expressed, in particular in fetal  
CC kidney, lung and liver, and in adult testis and liver. Also  
CC expressed in peripheral blood leukocytes, colon and small  
CC intestine, ovary, prostate, thymus, spleen, pancreas, kidney,  
CC lung, placenta and heart.  
CC -1- SIMILARITY: Contains 3 TNFR-Cys repeats.  
CC -1- SIMILARITY: Contains 1 death domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AF029761; AAD03477.1; -  
DR EMBL: AF021232; AAC32765.1; -  
DR EMBL: AF021233; AAC32766.1; -  
DR EMBL: AF023849; AAC52053.1; -  
DR HSSP: O14763; ID4V.  
DR GeneW: HGNC:11907; TNFRSF10D.  
DR MIM: 603614; -  
DR InterPro: IPR001368; TNFR\_C6.  
DR Pfam: PF00020; TNFR\_C6; 2.  
DR SMART: SM00208; TNFR\_2.  
DR PROSITE: PS00652; TNFR\_NGFR\_1; 2.  
DR PROSITE: PS50050; TNFR\_NGFR\_2; 2.  
KW Receptor; Apoptosis; Glycoprotein; Transmembrane; Repeat; Signal;  
KW Polymorphism.  
FT SIGNAL 1 55  
FT CHAIN 56 386  
FT DOMAIN 56 211  
FT TNSMSEM 212 232  
FT DOMAIN 233 386  
FT REPEAT 58 97  
FT REPEAT 98 139  
FT REPEAT 140 180  
FT DOMAIN 340 366  
FT DOMAIN 225 228  
FT DOMAIN 246 250  
FT DISULFID 83 96  
FT DISULFID 99 115  
FT DISULFID 99 115

TUMOR NECROSIS FACTOR RECEPTOR  
SUPERFAMILY MEMBER 10D.  
EXTRACELLULAR (POTENTIAL).  
POTENTIAL.  
CYTOPLASMIC (POTENTIAL).  
TNFR-CYS 1.  
TNFR-CYS 2.  
TNFR-CYS 3.  
DEATH (TRUNCATED).  
POLY-VAL.  
POLY-CYS.  
BY SIMILARITY.  
BY SIMILARITY.

```

FT DISULFID 118 131 BY SIMILARITY.
FT DISULFID 121 139 BY SIMILARITY.
FT DISULFID 141 155 BY SIMILARITY.
FT DISULFID 158 172 BY SIMILARITY.
FT DISULFID 162 180 BY SIMILARITY.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 182 182 P -> S (IN TRAIL-R4-B). (POTENTIAL).
FT VARIANT 35 35 /FTID-VAR 011417.
FT VARIANT 310 310 S -> L (IN TRAIL-R4-B).
FT SEQUENCE 386 AA; 41823 MW; 77A93577CAAF2632 CRC64;
SQ SEQUENCE 386 AA; 41823 MW; 77A93577CAAF2632 CRC64;

Query Match
Best Local Similarity 34.5%; Score 59; DB 1; Length 386;
Matches 12; Conservative 8; Mismatches 6; Indels 2; Gaps 1;

OY 2 GIPCCPVHKKRLIVVVVIVVIVG 29
Db 204 GMLASPHYH--LIIVVIVVIVVIVG 229

RESULT 11
SNN_HUMAN STANDARD; PRT; 88 AA.
AC 075324;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Stannin (AG8_1).
GN SNN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98325608; PubMed=9657854;
RA Dejneka N.S., Polavarapu R., Deng X., Martin-Deleon P.A.,
RA Billingsley M.L.
"Chromosomal localization and characterization of the stannin (Snn) gene."
RT Mamm. Genome 9:556-564(1998).
RL [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Endothelial cells;
RX MEDLINE=99252096; PubMed=10233894;
RA Horrevorts A.J.G., Fontijn R.D., Van Zonneveld A.J., de Vries C.J.M.,
RA ten Cate J.W., Pannekoek H.;
"Vascular endothelial genes that are responsible for tumor necrosis factor-alpha in vitro are expressed in atherosclerotic lesions, including inhibitor of apoptosis protein-1, stannin, and two novel genes."
RT Blood 93:3418-3431(1999).
RL [3]
RN SEQUENCE FROM N.A.
RC TISSUE=Amygdala;
RA Anserge W., Winkler U., Mewes H.-W., Weil B., Wiemann S.;
Submitted (MAR-2000) to the EMBL/Genbank/DBD databases.
-1- FUNCTION: PLAYS A ROLE IN THE TOXIC EFFECTS OF ORGANOTINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF030196; AAC28427.1; -
DR EMBL; AF070673; AAC83231.1; -
DR EMBL; AL161976; CAB82314.1; -

```

```

DR PIR; T47139; T47139.
DR Genew; HGNC:11149; SNN.
DR MIM: 603032; -
DR CO; GO:0003628; P:response to abiotic stimulus; TAS.
DR CO; GO:0006950; P:response to stress; TAS.
FW Transmembrane.
FT TRANSMEM 11 31 POTENTIAL.
SQ SEQUENCE 88 AA; 9497 MW; 1E8DA73323C5D6DF CRC64;

Query Match
Best Local Similarity 32.7%; Score 56; DB 1;
Matches 9; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

OY 15 IVVVVVVIVVIVGALIMG 34
Db 12 VVTIVVIVVIVVIVVIVVIVG 31

RESULT 12
SNN_MOUSE STANDARD; PRT; 88 AA.
AC 088369;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Stannin.
GN SNN.
OS Mus musculus (Mouse), and
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090, 10116;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Mouse;
RX MEDLINE=98325608; PubMed=9657854;
RA Dejneka N.S., Polavarapu R., Deng X., Martin-Deleon P.A.,
RA Billingsley M.L.
"Chromosomal localization and characterization of the stannin (Snn) gene."
RT Mamm. Genome 9:556-564(1998).
RL [2]
RN SEQUENCE FROM N.A.
RP SPECIES=Rat;
RX MEDLINE=92342135; PubMed=1635553;
RA Togas S.M., Krady J.K., Billingsley M.L.;
"molecular neurotoxicology of trimethyltin: identification of stannin, a novel protein expressed in trimethyltin-sensitive cells."
RT Mol. Pharmacol. 42:44-56(1992).
RL [3]
RN TISSUE SPECIFICITY.
RC SPECIES=Rat;
RX MEDLINE=98075763; PubMed=9413842;
RA Dejneka N.S., Patanow C.M., Polavarapu R., Togas S.M., Krady J.K.,
RA Billingsley M.L.;
"localization and characterization of stannin: relationship to cellular sensitivity to organotin compounds."
RT Neurochem. Int. 31:801-815(1997).
-1- FUNCTION: PLAYS A ROLE IN THE TOXIC EFFECTS OF ORGANOTINS.
-1- TISSUE SPECIFICITY: HIGH LEVEL OF EXPRESSION IN SPLEEN, FOLLOWED BY BRAIN AND KIDNEY.
-1- INDUCTION: BY TRIMETHYLTIN (TMT). A TRIALKYL TIN COMPOUND WHICH IS A POTENT NEUROTOXIC AGENT THAT SELECTIVELY DAMAGES SPECIFIC BRAIN REGIONS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; AF030522; AAC28428.1; -.
DR MGD; MGI:1276549; Snn.
KM Transmembrane.
FT TRANSMEM 11
SQ SEQUENCE 88 AA; 9501 MW; EBBDA73323D816C5 CRC64;

Query Match
Best Local Similarity 45.0%; Pred. No. 5.9;
Matches 9; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

OY 15 IVVVVVVVVVVIVGALLMG 34
Db 12 VVVVVILLVIAALGALLIG 31

RESULT 13
ID YIF4_YEAST STANDARD; PRT; 105 AA.
AC P40524;
DT 01-FEB-1995 (Rel. 31, last sequence update)
DT 01-FEB-1995 (Rel. 31, last sequence update)
DT 15-SEP-2003 (Rel. 42, last annotation update)
DE Hypothetical 12.0 kDa protein in SNPI-GPPI intergenic region.
GN YIL054W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX PubMed=9169870;
RA Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
RA Harris D.E., Holsen T., Hunt S., Jagels S., Jones M., Lye G.,
RA Mould S., O'Neill C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
RA Skelton J., Smith V., Walsh S., Whitehead S., Bartell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX."
RL Nature 387:84-87(1997).

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z38060; CA86168.1; -.
DR PIR; S48425; S48425.
DR SGD; S0001316; FYV2.
KM Hypothetical protein; Transmembrane.
FT TRANSMEM 49
FT TRANSMEM 71
SQ SEQUENCE 105 AA; 12045 MW; 7FBB73A4613CA181 CRC64;

Query Match
Best Local Similarity 23.3%; Pred. No. 6.7;
Matches 7; Conservative 11; Mismatches 12; Indels 0; Gaps 0;

OY 6 CPVHLKRLIVVVVIVVIVGALLMGL 35
Db 47 CPVFFMGIIITITVTLAIINIIFLT 76

RESULT 14
ID STX3_MOUSE STANDARD; PRT; 289 AA.
AC O64704;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 15-SEP-2003 (Rel. 42, last annotation update)

```

```

DE Syntaxin 3.
GN STX3A OR STX3.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN (1)
RP SEQUENCE FROM N.A. (ISOFORMS 3A; 3B; 3C; 3D; 3D-1 AND 3D-2).
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=95321968; PubMed=7598732;
RA Ibaraki K., Horikawa H.P.M., Morita T., Mori H., Sakimura K.,
RA Mishina M., Saito H., Abe T.;
RT "Identification of four different forms of syntaxin 3."
RL Biochem. Biophys. Res. Commun. 211:997-1005(1995).

CC -1- FUNCTION: POTENTIALLY INVOLVED IN DOCKING OF SYNAPTIC VESICLES AT
CC PRESYNAPTIC ACTIVE ZONES.
CC -1- SUBCELLULAR LOCATION: Type IV membrane protein (Potential).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=6;
CC Name=3A;
CC IsoId=Q64704-1; Sequence=Displayed;
CC Name=3B;
CC IsoId=Q64704-2; Sequence=VSP_006345;
CC Name=3C;
CC IsoId=Q64704-3; Sequence=VSP_006341, VSP_006346;
CC Name=3D;
CC IsoId=Q64704-4; Sequence=VSP_006344;
CC Name=3D-1;
CC IsoId=Q64704-5; Sequence=VSP_006342;
CC Name=3D-2;
CC IsoId=Q64704-6; Sequence=VSP_006343;
CC -1- SIMILARITY: BELONGS TO THE SYNTAXIN/EPIMORPHIN FAMILY.
CC -1- SIMILARITY: Contains 1 t-SNARE coiled-coil homology domain.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D29797; BAA06180.1; -.
DR EMBL; D29798; BAA06181.1; ALT_TERM.
DR EMBL; D29799; BAA06182.1; -.
DR EMBL; D29800; BAA06183.1; -.
DR EMBL; D38375; BAA07454.1; ALT_SEQ.
DR PIR; I60170; I60170.
DR PIR; I83197; I83197.
DR PIR; I83198; I83198.
DR HSSP; P33851; IBR0.
DR MGD; MGI:101077; Stx3.
DR InterPro; IPR006012; Syntaxin.
DR InterPro; IPR006011; Syntaxin_N.
DR InterPro; IPR000727; t-SNARE.
DR Pfam; PF00804; Syntaxin; 1.
DR SMART; SM00503; Sytn; 1.
DR SMART; SM00397; t-SNARE; 1.
DR PROSITE; PS00914; SYNTAXIN; 1.
DR PROSITE; PS50192; t-SNARE; 1.
KM Neurotransmitter transport; Coiled coil; Transmembrane;
KM Alternative splicing.
FT TRANSMEM 1
FT TRANSMEM 264
FT TRANSMEM 283
FT DOMAIN 284
FT DOMAIN 32
FT DOMAIN 191
FT VARSPPLIC 39
FT VARSPPLIC 72
FT VARSPPLIC 73
FT VARSPPLIC 86

CYTOPLASMIC (POTENTIAL).
ANCHOR FOR TYPE IV MEMBRANE PROTEIN
(POTENTIAL).
EXTRACELLULAR (POTENTIAL).
COILED COIL (POTENTIAL).
t-SNARE COILED-COIL HOMOLGY.
IESTRLINKISHVEAEAKVYIIISAPPEK -> NPH
GLISYLKRSSEH (in isoform 3C).
/FTId=VSP_006341.
TKDDLEQLTEIK -> LPWNPISPEIELT (in
isoform 3D-1).

```

QY	DB	Seq1	Seq2	Score	DB 1	Length	Indels	Gaps
QY	11	KRLIVVVVVVLIIVVIGALIMGL 35		32.2%	Score 55;	289;		
		Best Local Similarity 44.0%;			Pred. No. 18;			
Db	262	KKLIIIIIVVVVVLGIL--ALITGL 284			Matches 11; Conservative 11; Mismatches 1; Indels 2; Gaps 1			

ID	STX3_RAT	STANDARD:	PRT:	289 AA.
AC	008849;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Syntaxin 3.			
GN	STX3A OR STX3			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RA	SEQUENCE FROM N.A.			
RX	MEDLINE=93386759; PubMed=7690687;			
RA	Bennett M.K., Garcia-Arteras J.E., Elferink L.A., Peterson K.E.,			
RA	Fleming A.M., Hazuka C.D., Scheller R.H.;			
RT	"The syntaxin family of vesicular transport receptors.;"			
RL	Cell 74:863-873(1993)			
CC	-1- FUNCTION: POTENTIALLY INVOLVED IN DOCKING OF SYNAPTIC VESICLES			
CC	-1- AT PRESYNAPTIC ACTIVE ZONES.			
CC	-1- SUBCELLULAR LOCATION: Type IV membrane protein (Potential).			
CC	-1- TISSUE SPECIFICITY: HEART, SPLEEN, LUNG, AND KIDNEY.			
CC	-1- SIMILARITY: BELONGS TO THE SYNTAXIN/EPIMORPHIN FAMILY.			
CC	-1- SIMILARITY: Contains 1 t-SNARE coiled-coil homology domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by, and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; L20820; AAA03045.1; -.			
DR	PIR; D48213; D48213.			
DR	HSP; P32851; 18R0.			
DR	IncePro; IPR006012; Syntaxin.			
DR	IncePro; IPR006011; Syntaxin.N.			
DR	IncePro; IPR00727; T_SNARE.			
DR	Pfam; PF00804; Syntaxin.1.			
DR	SMART; SM00503; SYN; 1.			
DR	SMART; SM00397; t-SNARE; 1.			
DR	PROSITE; PS00914; SYNTAXIN.1.			
DR	PROSITE; PS50192; T_SNARE; 1.			

KM	Neurotransmitter transport; Coiled coil; Transmembrane.
FT	DOMAIN 1 263
FT	TRANSMEM 264 284
FT	
FT	DOMAIN 285 289
FT	DOMAIN 32 111
FT	DOMAIN 191 253
SQ	SEQUENCE 289 AA; 33257 MW; 4A8EAFD4684EEB6F CRC64;
<hr/>	
Query March	31.6%; Score 54; DB 1; Length 289;
Best Local Similarity	40.0%; Pred. No. 23;
Matches 10; Conservative 12; Mismatches	1; Indels 2; Gaps 1
Oy	11 KRLIIIVVVVLIIVIIGALLMGCL_35  ::  ::  ::  ::  ::   Db 262 KKLLIIVIIVVLGLL-ALITGL_284

Search completed: December 3, 2003, 15:50:33  
Job time : 7.19048 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 15:44:20 / Search time 25 Seconds

(without alignments)  
361.274 Million cell updates/sec

Title: US-09-788-308D-3

Sequence: 1 FGIPCCPVHLKRLIVVVVVIVGALLMGL 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeop:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	162	94.7	86	11 Q99MI3	Q99MI3 sigmodon hi
2	149	87.1	187	6 Q9TT04	Q9TT04 ovine aries
3	149	87.1	190	6 Q9N276	Q9N276 ovine aries
4	149	87.1	190	6 Q9BDX5	Q9BDX5 bovine taurus
5	66	38.6	345	7 Q8NM52	Q8NM52 oncorhynchus
6	66	38.6	770	5 Q9U9P2	Q9U9P2 hydra attenu
7	62	36.3	113	11 Q8C8J2	Q8C8J2 mus musculus
8	62	36.3	193	16 Q9X056	Q9X056 thermotoga
9	60	33.1	429	17 Q9Y91	Q9Y91 aetopyrum p
10	59.5	33.8	432	12 Q9DMH7	Q9DMH7 rat cytochrome
11	59	34.5	181	4 Q96N68	Q96N68 homo sapien
12	58	33.9	165	16 Q9PD60	Q9PD60 xylorella fas
13	58	33.9	315	16 Q9K32	Q9K32 bacillus ha
14	58	33.9	323	12 Q8CQ72	Q8CQ72 camelipox vi
15	58	33.9	334	16 Q98817	Q98817 rhizobium l
16	58	33.9	349	12 Q8V2X7	Q8V2X7 camelipox vi

17	58	33.9	349	16 Q8CSK3	Q8CSK3 staphylococ
18	58	33.9	415	13 Q8JG69	Q8JG69 brachydantio
19	57	33.3	166	16 Q9PP09	Q9PP09 ureaplasma
20	57	33.3	171	9 Q9MC12	Q9MC12 bacterioph
21	57	33.3	440	17 Q8TGL	Q8TGL pyrobaculum
22	56.5	32.7	451	8 Q9T7K5	Q9T7K5 crassostrea
23	56	32.7	534	16 Q9AB19	Q9AB19 caulobacter
24	56	32.7	745	10 Q8M0U2	Q8M0U2 sorghum bic
25	55.5	32.5	223	11 Q921A2	Q921A2 mus musculu
26	55.5	32.5	324	11 Q61735	Q61735 mus musculu
27	55.5	32.5	432	2 Q93UX6	Q93UX6 agrobacteri
28	55	32.2	29	16 Q9KCV2	Q9KCV2 bacillus ha
29	55	32.2	129	3 Q05899	Q05899 saccharomye
30	55	32.2	345	16 Q82059	Q82059 anabaena sp
31	55	32.2	619	5 Q01610	Q01610 caenorhabd
32	55	32.2	646	17 Q96ZB9	Q96ZB9 bulfoleobus
33	54.5	31.9	58	16 Q8CL58	Q8CL58 yersinia pe
34	54.5	31.9	156	16 Q9RK25	Q9RK25 streptomyce
35	54.5	31.9	229	8 Q9MLX7	Q9MLX7 sceloporus
36	54.5	31.9	268	5 Q17683	Q17683 caenorhabd
37	54	31.6	130	16 Q9S265	Q9S265 streptomyce
38	54	31.6	149	5 Q22745	Q22745 caenorhabd
39	54	31.6	151	16 Q9X854	Q9X854 streptomyce
40	54	31.6	183	12 Q65272	Q65272 african swi
41	54	31.6	183	12 Q65194	Q65194 african swi
42	54	31.6	183	16 Q25489	Q25489 helicobacte
43	54	31.6	185	12 Q65271	Q65271 african swi
44	54	31.6	195	16 Q8CQ23	Q8CQ23 staphylococ
45	54	31.6	288	16 Q9J553	Q9J553 chlamydia p

## ALIGNMENTS

Q99MI3	PRELIMINARY;	PRT;	86 AA
ID Q99MI3			
AC Q99MI3			
DT 01-JUN-2001 (TREMBLrel. 17, Created)			
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE Surfactant protein C (Fragment).			
OS Sigmodon hispidus (Hispid cotton rat).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Sigmodontinae;			
OC Sigmodon.			
OX NCBI_TaxID=42415;			
RN [1]			
RP SEQUENCE FROM N.A.			
RA Bennett J., Guzowski J., Langley R., Pletneva L., Ottoloni M.,			
RT "Influenza A Infection Increases Surfactant Protein C Gene Expression			
RT in the Cotton Rat (Sigmodon hispidus)."			
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.			
RN [2]			
RP SEQUENCE FROM N.A.			
RA Baccot S., Langley R.,			
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.			
DR EMBL; AF339111; AAK2618.1; -			
DR InterPro; IPR001729; Pulm surfact AP.			
DR ProDom; PD009591; Pulm surfact AP.1.			
DR PROSITE; PS00341; SURFACT_PALMTOTYL; 1.			
FT NON_TER	1		
FT NON_TER	1		
SQ	SEQUENCE	86 AA; 9333 MW; E85F946594639191 CRC64;	
Query Match	94.7%;	Score 162; DB 11; Length 86;	
Best Local Similarity	94.3%;	Pred. No. 1.1e-11;	
Matches	33; Conservative	1; Mismatches	1; Indels
Qy	1 FGIPCCPVHLKRLIVVVVVIVVIVGALLMGL 35		
Db	6 FGIPCCPVHLKRLIVVVVVIVVIVGALLMGL 40		

## RESULT 2

Q9TT04 PRELIMINARY; PRT; 187 AA.  
AC Q9TT04;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Pulmonary surfactant protein C (fragment).  
GN SPCS.  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Caprinae; Ovis.  
NCBI\_TaxID=9940;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Braems G.A., Yao L.-J., Inchley K., Brickenden A., Han V.K.M.,  
RA Grolla A., Challis J.R.G., Possmayer F.;  
RT "CDNA for ovine surfactant proteins: application in studies on fetal  
lung growth and maturation following prolonged hypoxemia."  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF211858; AAF18997.1; -  
DR InterPro; IPR001729; Pulm surfact AP.  
DR ProDom; PD009591; Pulm surfact AP. 1.  
DR PROSITE; PS00341; SURFACT\_PALMITOYL. 1.  
FT NON\_TER 1  
SQ SEQUENCE 187 AA; 20090 MW; 39B42072E7B809DD CRC64;

Query Match 87.1%; Score 149; DB 6; Length 187;  
Best Local Similarity 90.9%; Pred. No. 6.2e-10;  
Matches 30; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 IPCCPVHLKRLIVVVVVVIVGALLMGL 35  
Db 23 IPCCPVNIKRLIVVVVVVIVGALLMGL 55

## RESULT 3

Q9N276 PRELIMINARY; PRT; 190 AA.  
AC Q9N276;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Pulmonary surfactant-associated protein C.  
GN SP-C.  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Caprinae; Ovis.  
NCBI\_TaxID=9940;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA TISSUE=Lung;  
RC MEDLINE=20215263; PubMed=10749754;  
RX pieteschmann S.M., Pison U.;  
RT "CDNA cloning of ovine pulmonary SP-A, SP-B, and SP-C: isolation of  
two different sequences for SP-B."  
RL Am. J. Physiol. 278:L765-L782(2000).  
DR EMBL; AF076634; AAF31149.1; -  
DR InterPro; IPR001729; Pulm surfact AP.  
DR ProDom; PD009591; Pulm surfact AP. 1.  
DR PROSITE; PS00341; SURFACT\_PALMITOYL. 1.  
SQ SEQUENCE 190 AA; 20337 MW; B9F6440990AF31BD CRC64;

Query Match 87.1%; Score 149; DB 6; Length 190;  
Best Local Similarity 90.9%; Pred. No. 6.3e-10;  
Matches 30; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 IPCCPVHLKRLIVVVVVVIVGALLMGL 35  
Db 26 IPCCPVNIKRLIVVVVVVIVGALLMGL 58

## RESULT 4

Q9BDX5 PRELIMINARY; PRT; 190 AA.  
AC Q9BDX5;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Pulmonary surfactant-associated protein C proSP-C.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lung;  
RA Daniels F., Vandebussche G., Ruyschaert J.-M.;  
RT "CDNA cloning of bovine (Belgian White and Blue breed) pulmonary  
surfactant-associated protein C."  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF250035; AAK28414.1; -  
DR InterPro; IPR001729; Pulm surfact AP.  
DR ProDom; PD009591; Pulm surfact AP. 1.  
DR PROSITE; PS00341; SURFACT\_PALMITOYL. 1.  
SQ SEQUENCE 190 AA; 20440 MW; BD9508798ABD2F43 CRC64;

Query Match 87.1%; Score 149; DB 6; Length 190;  
Best Local Similarity 90.9%; Pred. No. 6.3e-10;  
Matches 30; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 IPCCPVHLKRLIVVVVVVIVGALLMGL 35  
Db 26 IPCCPVNIKRLIVVVVVVIVGALLMGL 58

## RESULT 5

Q8WM52 PRELIMINARY; PRT; 345 AA.  
AC Q8WM52;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE MHC class I antigen (fragment).  
GN OMWY-071.  
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
OC Proteocephali; Salmoniformes; Salmonidae; Oncorhynchus.  
NCBI\_TaxID=8022;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Shum B.P., Mason P.M., Magor K.E., Flodin L.R., Stet R., Parham P.;  
RT "Structures of Two Major Histocompatibility Complex (MHC) Class I  
Genes of the Rainbow Trout (Oncorhynchus mykiss)."  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
CC - FUNCTION INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE  
CC IMMUNE SYSTEM (BY SIMILARITY).  
CC - SUBUNIT DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-  
CC MICROGLOBULIN) (BY SIMILARITY).  
DR EMBL; AY071854; AA60588.1; -  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003597; Ig\_c1.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR001039; MHC\_I.  
DR Pfam; PF00047; Ig\_1.  
DR Pfam; PF00129; MHC\_I. 1.  
DR ProDom; PD000050; MHC\_I. 1.  
DR SMART; SM00407; IgC1. 1.  
DR PROSITE; PS50835; IG\_LIKE. 1.  
KM Glycoprotein; Transmembrane.  
FT NON\_TER 1

SO SEQUENCE 345 AA; 39393 MW; 16AD5C77C303D167 CRC64;

Query Match 38.6%; Score 66; DB 7; Length 345;  
Best Local Similarity 61.9%; Pred. No. 3.7;  
Matches 13; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 13 LLIIVVVVVVLIIVVIGALLM 33  
||:|||||:|||||:|:  
DB 294 LLIIVVVVVVVVVVVVVVVVV 314

## RESULT 6

Q9U9P2 PRELIMINARY; PRT; 770 AA.

AC Q9U9P2; 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
GN Endothelin converting enzyme.  
OS Hydra attenuata (Hydra) (Hydra vulgaris).  
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroidea; Anthomedusae;  
OC Hydridae; Hydra.  
OX NCBI\_TaxID=6087;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21185982; PubMed=11290299;  
RA Zhang J., Leontovich A., Sarrae M.P., Jr.:  
RT "Molecular and functional evidence for early divergence of an  
RT endothelin-like system during metazoan evolution: analysis of the  
RT Cnidarian, hydra."  
RL Development 128:1607-1615(2001).  
DR EMBL: AF162671; AAD46624.1; -.  
DR HSSP; P08473; IDMT.  
DR InterPro; IPR000718; Peptidase\_M13.  
DR InterPro; IPR006025; Zn\_MTPeptidase.  
DR Pfam; PF01431; Peptidase\_M13; 1.  
DR PRINTS; PRO0786; NEPRILYSIN.  
DR PROSITE; PS00142; ZINC PROTEASE; 1.  
SO SEQUENCE 770 AA; 88599 MW; D44C822F4BEE896D CRC64;

Query Match 38.6%; Score 66; DB 5; Length 770;  
Best Local Similarity 40.7%; Pred. No. 6.7;  
Matches 11; Conservative 12; Mismatches 4; Indels 0; Gaps 0;

QY 9 HLKRLIIVVVVLIIVVIGALLMGL 35  
||:|||||:|||||:|:  
DB 39 HLRLIIVVVVLIIVVLIIVVIGALLMGL 65

## RESULT 7

Q8C8J2 PRELIMINARY; PRT; 113 AA.

AC Q8C8J2; 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
GN Weakly similar to probable glycosyl transferase.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA "The RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
DR EMBL; AK046933; BAC12922.1; -.  
SO SEQUENCE 113 AA; 12876 MW; 03186A70C3F74A12 CRC64;

Query Match 36.3%; Score 62; DB 11; Length 113;  
Best Local Similarity 59.1%; Pred. No. 4.7;  
Matches 13; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 13 LLIIVVVVVVLIIVVIGALLMGL 34  
||:|||||:|||||:|:  
DB 37 LLIIVVVVVVVVVVVVVVVVV 58

## RESULT 8

Q9X056 PRELIMINARY; PRT; 193 AA.

AC Q9X056; 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DE LEMA protein.  
GN TM0961.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109;  
RX MEDLINE=99287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,  
RA Halt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
RA Stewart A.M., Cotton M.G., Pratt M.S., Phillips C.A., Richardson D.,  
RA Heidelberg J., Sutton G.D., Fleischmann R.D., Eisen J.A., White O.,  
RA Salzberg S.L., Smith H.O., Venter J.C., Frazer C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
RT genome sequence of Thermotoga maritima."  
RL Nature 399:323-329(1999).  
DR EMBL; AB001759; AAD36040.1; -.  
DR TIGR; TM0961; -.  
KW Complete proteome.  
SO SEQUENCE 193 AA; 21635 MW; F2C04F81DB9F257 CRC64;

Query Match 36.3%; Score 62; DB 16; Length 193;  
Best Local Similarity 50.0%; Pred. No. 7;  
Matches 11; Conservative 9; Mismatches 2; Indels 0; Gaps 0;

QY 8 VHLKRLIIVVVVLIIVVIG 29  
||:|||||:|||||:|:  
DB 5 IRVKGRLIIVVLIIVVIG 26

## RESULT 9

Q9YG91 PRELIMINARY; PRT; 538 AA.

AC Q9YG91; 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
GN Hypothetical protein APE0010.  
GN APE0010.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcaceae;  
OC Desulfurococcaceae; Aeropyrum.  
OX NCBI\_TaxID=56636;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KI;  
RX MEDLINE=99110339; PubMed=10382966;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Hatake Y.,  
RA Jinno K., Takahashi M., Sekine M., Baba S.-I., Akai A., Kosugi H.,  
RA Hoozono A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
RA Yamazaki J., Kusuda N., Oguchi A., Aoki K.-I., Kubota K.,  
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic

RT crenarchaeon, Aeropyrum pernix K1."  
 RL DNA Res. 6:83-101(1999).  
 DR EMBL: AP000058; BAA78919.1; -  
 DR InterPro: IPR002293; RA/rel\_permease.  
 DR InterPro: IPR002934; NTP\_transf.  
 DR InterPro: IPR004841; Permease.  
 DR Pfam: PF00324; aa\_permeases; 1.  
 DR Pfam: PF01909; NTP\_transf.2; 1.  
 KM Hypothetical protein: Complete proteome.  
 SQ SEQUENCE 538 AA; 57642 MW; 9724088023BFEPS CRC64;

Query Match 35.1%; Score 60; DB 17; Length 538;  
 Best Local Similarity 40.6%; Pred. No. 26;  
 Matches 13; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

Qy 3 IPCCPVHLKRLIVVVVLIIVVIGALLMG 34  
 Db 121 INCCPREAYIALVVFVIALGLVVMGLAIMG 152

RESULT 10  
 Q9DWH7 PRELIMINARY; PRT; 429 AA.  
 AC Q9DWH7; 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Pr2.1.  
 GN R2.1.  
 OS Rat cytomegalovirus (strain Mastricht).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Muromegalovirus.  
 OX NCBI\_TaxID=79700;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mastricht;  
 RX MEDLINE=20366325; PubMed=10906222;  
 RA Vink C., Beuken E., Bruggeman C.A.;  
 RT "Complete DNA sequence of the rat cytomegalovirus genome."  
 RL J. Virol. 74:7656-7665(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mastricht;  
 RX MEDLINE=20473137; PubMed=11018281;  
 RA Gruijthuisen Y.K., Beuken E., Bruggeman C.A.;  
 RT "Rat cytomegalovirus R89 is a highly conserved gene which expresses a  
 RT spliced transcript."  
 RL Virus Res. 69:119-130(2000).  
 DR EMBL: AF232689; AAF99112.1; -  
 SQ SEQUENCE 429 AA; 41925 MW; 217520F5CACA7373 CRC64;

Query Match 34.8%; Score 59.5; DB 12; Length 429;  
 Best Local Similarity 48.1%; Pred. No. 25;  
 Matches 13; Conservative 9; Mismatches 4; Indels 1; Gaps 1;

Qy 2 GIPCCPVHLKRLIVVVVLIIVVIV 28  
 Db 330 GVP-VPLPRPVVVVVVVVVVVVVVV 355

RESULT 11  
 Q96N68 PRELIMINARY; PRT; 181 AA.  
 AC Q96N68; 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Hypothetical protein FLJ31338.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Mammary gland;  
 RA Teshiro H., Yamazaki M., Matanabe K., Kumagai A., Itakura S.,  
 RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,  
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
 RA Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Nagasawa M.,  
 RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,  
 RA Kwakami B., Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K.,  
 RA Iwaga T.;  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK055900; BAB71040.1; -  
 KM Hypothetical protein.  
 SQ SEQUENCE 181 AA; 19136 MW; C427404F1F954CAA CRC64;

Query Match 34.5%; Score 59; DB 4; Length 181;  
 Best Local Similarity 44.4%; Pred. No. 15;  
 Matches 12; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

Qy 4 PCPCPVHLKRLIVVVVLIIVVIVGA 30  
 Db 115 PCPCPVHQSILTVVCMCAKCVHVCVA 141

RESULT 12  
 Q9PD60 PRELIMINARY; PRT; 165 AA.  
 ID Q9PD60; 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE General secretory pathway protein G precursor.  
 GN Xp1519.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylella.  
 OX NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=985C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Agencio M.,  
 RA Alvaranga R., Alves L.M.C., Araya J.E., Baia J.M., Baptista C.S.,  
 RA Barrios M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Brites M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carer H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.B., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferrito J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frome M., Furian L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Honeisel J.D., Junqueira M.L., Kemper B.L., Kitajima J.P.,  
 RA Klieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsunuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nanti A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Paixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.B. Jr., de Sa R.G., Santelli R.V., Sasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Trufi D., Tsai S.M., Tsubako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Zeldanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa."  
 RL Nature 406:151-159(2000)

CC -1- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE  
 CC EXPORT OF PROTEINS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE PULG/OUTG/XPSG/EXEG/XCPT FAMILY.



DR EMBL; AE003982; AAF84328.1; -.  
 DR InterPro; IPR000983; Bac\_GSPG.  
 DR InterPro; IPR001120; Prok N methyltn.  
 DR PRINTS; PR00813; BACTERIALGSG.  
 DR PROSITE; PS00409; PROKAR\_NTER\_METHYL; 1.  
 KW Methylation; Transport; Complete proteome.  
 SQ SEQUENCE 165 AA; 17543 MW; 5480988B1650AD9F CRC64;

Query Match 33.9%; Score 58; DB 16; Length 165;  
 Best Local Similarity 34.2%; Pred. No. 18;  
 Matches 13; Conservative 11; Mismatches 8; Indels 6; Gaps 2;

Qy 3 IPCCPVHLKRLIIVVVVLI--VVIVGALLMG 34  
 Db 28 ITCSPAMQKAGMSLEIIIVIVIGVIAFVGSRLVG 65

RESULT 13  
 Q9KE32 PRELIMINARY; PRT; 315 AA.  
 AC Q9KE32;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Ferichrome ABC transporter (Permease).  
 GN BHI026.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxId=6665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 DR EMBL; AP001510; BAB04745.1; -.  
 DR InterPro; IPR000522; FeccD.  
 DR Pfam; PF01032; FeccD; 1.  
 DR Prodom; PD001557; FeccD; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 315 AA; 35224 MW; 977F95BD0751C89 CRC64;

Query Match 33.9%; Score 58; DB 16; Length 315;  
 Best Local Similarity 45.7%; Pred. No. 29;  
 Matches 16; Conservative 6; Mismatches 11; Indels 2; Gaps 2;

Qy 2 GIPCCPVHLKRLIIVVVVLI--VVIVGALLMG 35  
 Db 212 GIPYDSV-IKRLIIVVAFISIALVGBITFLGL 245

RESULT 14  
 Q80Q72 PRELIMINARY; PRT; 323 AA.  
 AC Q80Q72;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE CPM40L.  
 GN CPM40L.  
 OS Camel痘 virus (strain CP-1).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Orthopoxvirus.  
 OX NCBI\_TaxId=203174;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CMS;  
 RX PubMed=11907336;  
 RX Gubser C., Smith G.L.;

RT "The sequence of camel痘 virus shows it is most closely related to  
 RT variola virus, the cause of smallpox.";  
 RL J. Gen. Virol. 83:855-872(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CMS;  
 RA Gubser C., Smith G.L.;  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY009089; ANG37499.1; -.  
 SQ SEQUENCE 323 AA; 36824 MW; 767B4347D5AA9527 CRC64;

Query Match 33.9%; Score 58; DB 12; Length 323;  
 Best Local Similarity 25.8%; Pred. No. 30;  
 Matches 8; Conservative 16; Mismatches 3; Indels 4; Gaps 1;

Qy 7 PHLKRLIIVVVV---VLIIVIVGALLMG 33  
 Db 281 PMNVNLIIVIVLITLIIIIIVIVIAIYM 311

RESULT 15  
 Q98817 PRELIMINARY; PRT; 334 AA.  
 AC Q98817;  
 DT 01-OCT-2001 (TREMBLrel. 18, Created)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical protein mlr6726.  
 GN MLR6726.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OX NCBI\_TaxId=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WAF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Matenabe A., Ideesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti.";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL; AP003010; BABS2963.1; -.  
 DR InterPro; IPR002549; UPF0118.  
 DR Pfam; PF01594; UPF0118; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 334 AA; 36447 MW; 0E4D8323A594B2F6 CRC64;

Query Match 33.9%; Score 58; DB 16; Length 334;  
 Best Local Similarity 42.3%; Pred. No. 31;  
 Matches 11; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

Qy 10 LKRLIIVVVVLIIVIVGALLMG 35  
 Db 49 LAALITLIVLVIVGSAIVMGL 74

Search completed: December 3, 2003, 15:52:34  
 Job time : 27 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 3, 2003, 15:40:35 ; Search time 22.2041 Seconds  
(without alignments)  
171.565 Million cell updates/sec

Title: US-09-788-308D-4

Perfect score: 141  
Sequence: 1 FFPVHLKRGCGGCGGCGGCGG 24

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.GeneSeq 130un03.\*  
1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	74.5	155	22	ABBS5265
2	103.5	72.7	485	22	ABBS5608
3	101	71.6	81	23	ABBS7205
4	101	71.6	81	23	AAE20135
5	101	71.6	100	23	ABBS76203
6	101	71.6	100	23	AAE20133
7	101	71.6	100	24	ABG74445
8	101	71.6	304	22	AABS9556
9	101	71.6	304	23	ABP43856

10	101	71.6	520	22	AAAS2318
11	101	71.6	520	22	AAAG7350
12	101	71.6	520	22	AAAB9336
13	101	71.6	688	21	AAV57393
14	99	70.2	50	18	AAW13632
15	99	70.2	50	18	AAW13633
16	99	70.2	74	22	ABBS6951
17	99	70.2	282	24	ABRA41039
18	99	70.2	485	21	AAV69385
19	99	70.2	485	21	AAV69392
20	99	70.2	495	22	AAW78993
21	99	70.2	514	22	AAW79977
22	98	69.5	33	23	AAE28394
23	98	69.5	638	22	ABBS7747
24	98	69.5	2783	13	AAAR23962
25	98	69.5	2783	13	AAAR23963
26	98	69.5	2783	13	AAAB82946
27	97	68.8	94	22	AAAG71269
28	97	68.8	94	22	ABG63450
29	97	68.8	416	23	ABBS7171
30	96	68.1	19	21	AAV83221
31	96	68.1	20	22	AAAB6491
32	96	68.1	20	22	AAE28377
33	96	68.1	21	22	AAAB69492
34	96	68.1	23	20	AAW67767
35	96	68.1	24	19	AAW79139
36	96	68.1	56	23	ABP29010
37	96	68.1	101	22	ABG47671
38	96	68.1	101	22	ABBS27649
39	96	68.1	101	22	ABBS32819
40	96	68.1	101	22	ABBS18301
41	96	68.1	101	22	AAAS6026
42	96	68.1	101	22	AAAS6005
43	96	68.1	101	22	AAW13874
44	96	68.1	101	22	AAW26281
45	96	68.1	101	22	AAW01617

#### ALIGNMENTS

RESULT 1  
ABBS5265  
ID ABBS5265 standard; Protein; 155 AA.  
XX  
AC ABBS5265;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 22587.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
XX  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PMD, Myers BW;  
XX  
DR WPI: 2001-656860/75.  
XX  
DR N-PSDB; ABL09368.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more

Murine WASP protei  
Amino acid sequenc  
Murine WASP protei  
Mouse nerve differ  
Apolipoprotein (a)  
Apolipoprotein (a)  
Drosophila melanog  
Human MAP kinase c  
Amino acid sequenc  
Amino acid sequenc  
Human protein SEQ  
Human protein SEQ  
NLS based peptide  
Drosophila melanog  
APP-1. Homo sapi  
APP-1 (Ala 2460 Va  
Human androgen rec  
Human gene 1-encod  
Human albumin fusi  
Mouse ischaemic co  
Peptide linker use  
Synthetic HAV pept  
Peptide #3 used in  
Synthetic HAV pept  
Platelet glycoprot  
GG24 Gly-ala inner  
Streptococcus poly  
Human liver peptid  
Human Peptide #300  
Peptide #325 encod  
Protein #300 encod  
Human brain expres  
Human bone marrow  
Peptide #308 encod  
Peptide #318 encod  
Peptide #299 encod

PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 22587; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 155 AA;  
QY  
Query Match 74.5%; Score 105; DB 22; Length 155;  
Best Local Similarity 81.8%; Pred. No. 6.9e-06;  
Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
DB 3 PVHLKRGGGGGGGGGGGGGGGG 24  
49 PVYPKEGGGGGGGGGGGGGGG 70  
RESULT 2  
ABB59608  
ID ABB59608 standard; Protein; 485 AA.  
XX  
AC ABB59608;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 5616.  
XX  
KM Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
XX  
PT 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
PI Venter JC, Adams M, Li FWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL03711.  
XX  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX  
PS Disclosure; SEQ ID NO 5616; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins  
XX (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 485 AA;  
QY  
Query Match 72.7%; Score 102.5; DB 22; Length 485;  
Best Local Similarity 81.8%; Pred. No. 3.9e-05;  
Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;  
DB 3 PVHLKRGGGGGGGGGGGGGGGG 24  
395 PIH---GGGGGGGGGGGGGGGGG 413  
RESULT 3  
ABB76205  
ID ABB76205 standard; Protein; 81 AA.  
XX  
AC ABB76205;  
XX  
DT 09-AUG-2002 (first entry)  
XX  
DE Lantibody display peptide suna-PG20-SL.  
XX  
KM Lantibody; sublancin; subtilin.  
XX  
OS Bacillus subtilis.  
XX Synthetic.  
XX  
FH Key Location/Qualifiers  
XX  
FT Protein 1..38  
FT /note= "sublancin mature protein"  
FT Disulfide-bond 7..36  
FT Disulfide-bond 14..29  
FT Modified-site 16  
FT /note= "dehydroalanine"  
FT Modified-site 19  
FT /note= "Aba"  
FT Misc-difference 19  
FT /note= "linked to residue 22 via sulfide bond to  
FT form cyclic peptide"  
FT Misc-difference 22  
FT /note= "linked to residue 9 via sulfide bond to  
FT form cyclic peptide"  
FT Peptide 38..57  
FT /note= "polylglycine linker"  
FT Peptide 58..81  
FT /note= "subtilin leader"  
XX  
PN US2002052005-A1.  
XX  
XX 02-MAY-2002.  
XX  
XX 29-JUN-2001; 2001US-0893499.  
XX  
XX 29-JUN-2000; 2000US-215449P.  
XX  
XX (UTMA-) UNIV MARYLAND BALTIMORE.  
XX  
XX Hansen JN;  
XX  
XX WPI; 2002-415206/44.  
XX  
XX  
PT Detecting binding of a target molecule to a lantibody display peptide,  
PT useful for designing new lantibodies directed towards nucleophilic  
PT targets, involves expressing the peptide in a host cell -  
XX  
XX Example 4; Fig 8; 20pp; English.  
XX  
XX The present sequence is the protein sequence of suna-PG20-SL  
XX comprising the Bacillus subtilis sublancin mature protein, a  
XX 20-residue polylglycine spacer and the subtilin leader segment.

CC This lantibody display peptide was expressed from B. subtilis 168.  
CC The present invention relates to the construction and screening of  
CC a subclonin-derived lantibody display library in a strain of B.  
CC subtilis. A subclonin peptide variant (lantibody) having a spacer  
CC region and a subtilin leader peptide provides a form of subclonin  
CC that is retained in the cell wall instead of being released to the  
CC medium. Methods for generating a lantibody display library and for  
CC identifying new target molecules are provided.

XX Sequence 81 AA;

Query Match 71.6%; Score 101; DB 23; Length 81;  
Best Local Similarity 100.0%; Pred. No. 1.2e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 RCGGGGGGGGGGGGGG 24  
DB 37 RCGGGGGGGGGGGGGG 53

## RESULT 4

AAE20135 standard; Protein; 81 AA.

AC AAE20135;

DT 18-JUN-2002 (first entry)

XX Bacillus subtilis subclonin derived-lantibody display peptide.

XX Lantibody display peptide; lantibiotic peptide; spacer; subtilin leader;  
XX target molecule identification; sunA; subclonin; chimeric.

OS Chimeric - Bacillus subtilis.

OS Chimeric - Unidentified.

XX Key Location/Qualifiers

FT Region 1..37 /note= "Mature subclonin"

FT Disulfide-bond 7..36

FT Disulfide-bond 14..29

FT Modified-site 16 /note= "Dehydroalanine"

FT Modified-site 19 /note= "Aba: This residue is linked to Ala at position 22 by thioether linkage"

FT Modified-site 22 /note= "This residue is linked to Aba at position 19 by thioether linkage"

FT Region 38..57 /note= "Polylglycine spacer"

FT Region 58..81 /note= "Subtilin leader"

XX US2002019518-A1.

XX 14-FEB-2002.

XX 29-JUN-2001; 2001US-0893600.

XX 29-JUN-2000; 2000US-215449P.

XX (UYMA-) UNIV MARYLAND BALTIMORE.

XX Hansen JN;

XX WPI; 2002-239265/29.

XX Lantibiotic display peptide which allows functional antibody to be  
XX displayed on cell wall outer surface, has chimeric polypeptide having  
XX lantibiotic peptide, spacer attached to the peptide and subtilin leader  
XX segment -

PS Example 4; Fig 8; 20pp; English.

XX The invention relates to a subclonin-derived lantibody display peptide  
XX in a strain of Bacillus subtilis. Lantibody display peptide comprises  
XX a chimeric polypeptide having a lantibiotic peptide, an amino acid  
XX spacer attached to C-terminus of lantibiotic peptide and a subtilin  
XX leader segment attached to the spacer. The invention also provides  
XX methods for selecting lantibody display peptides having the ability  
XX to bind to nucleophilic groups on a target molecule. Kits containing  
XX a bacterial host cell expressing lantibody display peptides are used  
XX in the identification target molecules. The present sequence is  
XX lantibody display peptide which comprises a mature subclonin (sunA)  
XX peptide derived from Bacillus subtilis strain 168, a polylglycine  
XX spacer and a subtilin leader segment.

XX Sequence 81 AA;

Query Match 71.6%; Score 101; DB 23; Length 81;  
Best Local Similarity 100.0%; Pred. No. 1.2e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 RCGGGGGGGGGGGGGG 24  
DB 37 RCGGGGGGGGGGGGGG 53

## RESULT 5

ABB76203 standard; Protein; 100 AA.

AC ABB76203;

DT 09-AUG-2002 (first entry)

XX Lantibody display peptide sunA-PC20-SP.

XX Lantibody; subclonin; subtilin.

XX Bacillus subtilis.

OS Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..19 /label= "Signal peptide"

FT Peptide /note= "subclonin leader peptide"

FT Peptide 20..56 /label= "Mature protein"

FT Peptide 57..76 /note= "subclonin mature protein"

FT Peptide 77..100 /note= "polylglycine linker"

FT Peptide /label= "leader peptide"

FT Peptide /label= "subtilin leader"

XX US2002052005-A1.

XX 02-MAY-2002.

XX 29-JUN-2001; 2001US-0893499.

XX 29-JUN-2000; 2000US-215449P.

XX (UYMA-) UNIV MARYLAND BALTIMORE.

XX Hansen JN;

XX WPI; 2002-415206/44.

XX N-PSDB; ABL57235.

XX Detecting binding of a target molecule to a lantibody display peptide,  
XX useful for designing new lantibodies directed towards nucleophilic  
XX targets, involves expressing the peptide in a host cell -

PS Example 3; Fig 7; 20pp; English.  
XX  
CC The present sequence is the protein sequence of suna-PG20-SL  
CC comprising the Bacillus subtilis subtlancin leader peptide and mature  
CC protein, a 20-residue polyglycine spacer and the subtilin leader  
CC segment. This lantibody display peptide was expressed from B.  
CC subtilis 168 (see AB876205). The present invention relates to the  
CC construction and screening of a subtlancin-derived lantibody display  
CC library in a strain of B. subtilis. A subtlancin peptide variant  
CC (lantibody) having a spacer region and a subtilin leader peptide  
CC provides a form of subtlancin that is retained in the cell wall  
CC instead of being released to the medium. Methods for generating a  
CC lantibody display library and for identifying new target molecules  
CC are provided.  
XX  
SQ Sequence 100 AA;  
Query Match 71.6%; Score 101; DB 23; Length 100;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 8 RCGGGGGGGGGGGGGG 24  
Db 56 RCGGGGGGGGGGGGGG 72  
RESULT 6  
AAE20133 ID AAE20133 standard; Protein: 100 AA.  
XX  
AC AAE20133;  
XX  
DT 18-JUN-2002 (first entry)  
XX  
DE Bacillus subtilis Suna-PG20-SL peptide.  
XX  
KM Lantibody display peptide; lantibiotic peptide; spacer; subtilin leader;  
KM target molecule identification; chimeric; suna; subtlancin.  
XX  
OS Chimeric - Bacillus subtilis.  
XX  
OS Chimeric - Unidentified.  
XX  
FH Key Location/Qualifiers  
FT Region 1..19 /note= "Subtlancin leader"  
FT Region 20..26 /note= "Subtlancin prepeptide"  
FT Region 27..56 /note= "Mature Subtlancin"  
FT Region 57..76 /note= "Polyglycine spacer"  
FT Region 77..100 /note= "Subtilin leader"  
FT Region /note= "Subtilin leader"  
XX  
PN US2002019518-A1.  
XX  
PD 14-FEB-2002.  
XX  
PF 29-JUN-2001; 2001US-0893600.  
XX  
PR 29-JUN-2000; 2000US-215449P.  
XX  
PA (UYMA-) UNIV MARYLAND BALTIMORE.  
XX  
PI Hansen JN;  
XX  
DR WPI: 2002-239265/29.  
DR N-PSDB; AAD31956, AAD32009.  
XX  
PT Lantibiotic display peptide which allows functional antibody to be  
PT displaced on cell wall outer surface, has chimeric polypeptide having  
PT lantibiotic peptide, spacer attached to the peptide and subtilin leader  
PT segment -

XX  
PS Claim 4; Fig 7; 20pp; English.  
XX  
CC The invention relates to a subtlancin-derived lantibody display peptide  
CC in a strain of Bacillus subtilis. Lantibody display peptide comprises  
CC a chimeric polypeptide having a lantibiotic peptide, an amino acid  
CC spacer attached to C-terminus of lantibiotic peptide and a subtilin  
CC leader segment attached to the spacer. The invention also provides  
CC methods for selecting lantibody display peptides having the ability  
CC to bind to nucleophilic groups on a target molecule. Kits containing  
CC a bacterial host cell expressing lantibody display peptides are used  
CC in the identification target molecules. The present sequence is  
CC suna-PG20-SL chimeric peptide. This peptide comprises subtlancin (suna)  
CC leader, prepeptide and mature subtlancin from Bacillus subtilis  
CC strain 168, a polyglycine spacer and subtilin leader segment.  
XX  
SQ Sequence 100 AA;  
Query Match 71.6%; Score 101; DB 23; Length 100;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 8 RCGGGGGGGGGGGGGG 24  
Db 56 RCGGGGGGGGGGGGGG 72  
RESULT 7  
ABG74445 ID ABG74445 standard; Protein: 100 AA.  
XX  
AC ABG74445;  
XX  
DT 22-APR-2003 (first entry)  
XX  
DE Partial subtlancin protein.  
XX  
KM Subtlancin; affinity-tag labelled subtlancin peptide; decontamination;  
KM bacterial spore-infected area; spore-inhibition; industrial product;  
KM consumer product; food preservative; antimicrobial agent;  
KM bioremediation; environmental contamination; gram-positive spore former;  
KM Bacillus anthracis; anthrax.  
XX  
OS Bacillus subtilis.  
XX  
PN US2002165139-A1.  
XX  
PD 07-NOV-2002.  
XX  
PF 29-JUN-2001; 2001US-0894030.  
XX  
PR 29-JUN-2000; 2000US-215449P.  
XX  
PA (UYMA-) UNIV MARYLAND BALTIMORE.  
XX  
PI Hansen JN;  
XX  
DR WPI: 2003-238312/23.  
DR N-PSDB; ABX10836.  
XX  
PT Novel affinity-tag labeled subtlancin peptide, for decontaminating  
PT bacterial spore-infected area, has chimeric polypeptide having  
PT subtlancin peptide, spacer attached to C-terminus of peptide and tag  
PT attached to spacer -  
XX  
PS Disclosure; Page 8-9; 16pp; English.  
XX  
CC The invention describes an affinity-tag labelled subtlancin peptide (I)  
CC comprising a chimeric polypeptide having a subtlancin peptide, an amino  
CC acid spacer attached to the C-terminus of the subtlancin peptide, and an  
CC affinity tag attached to the spacer. (I) is useful for decontaminating a  
CC bacterial spore-infected area, by treating the infected area with a  
CC spore-inhibiting effective amount of (I). (I) is useful in medical,

Industrial and consumer products, as food preservative, and as an antimicrobial agent or decontaminating agent. (1) is useful for facilitating isolation of subcloning and in bioremediation of environmental contamination by gram-positive spore formers, such as *Bacillus anthracis* (the spores of which cause anthrax). (1) has specific inhibitory activity for spore outgrowth similar to native subcloning peptide. The affinity tag in (1) enables enhanced purification of peptide variant from sample preparation without affecting the intracellular processing of the subcloning peptide variant, expression by a host cell or its biological activity in secreted form. This is the amino acid sequence of a partial *Bacillus subtilis* subcloning protein used to create the affinity-tag labelled subcloning peptide of the invention.

SQ Sequence 100 AA;

Query Match 71.6%; Score 101; DB 24; Length 100;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 RGGGGGGGGGGGGGGG 24  
DB 56 RGGGGGGGGGGGGGGG 72

RESULT 8

ID AAB92556 standard; Protein; 304 AA.

AC AAB92556;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:10745.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EPI074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

PI WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

PT Claim 8; SEQ ID 10745; 2537bp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

XX full-length cDNAs defined in the specification. Where a primer set

XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary

XX to the complementary strand of a polynucleotide which comprises one of

XX the 5602 nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialized methods. AAH03166 to AAH13628 and AAH13632 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

SQ Sequence 304 AA;

Query Match 71.6%; Score 101; DB 22; Length 304;  
Best Local Similarity 89.5%; Pred. No. 3.9e-05;  
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 6 LKRGGGGGGGGGGGGGG 24  
DB 99 LRAGGGGGGGGGGGGGG 117

RESULT 9

ID ABP43856 standard; Protein; 304 AA.

AC ABP43856;

DT 26-FEB-2003 (first entry)

DE Zinc finger FEZL.

KW Neuroprotective; immunomodulator; cancer; chromosome 3;

KW cytoskeletal; anti-inflammatory; gene therapy; nutritional supplement;

KW wound; burn; ulcer; Alzheimer's disease; Huntington's disease;

KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;

KW vulnery.

OS Homo sapiens.

PN WO00231111-A2.

PD 18-APR-2002.

PF 11-OCT-2001; 2001WO-US27760.

PR 12-OCT-2000; 2000US-0687527.

PA (HYSE-) HYSEQ INC.

PI Tang YF, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

PI WPI; 2002-426278/45.

PT N-PSDB; AB061100.

PT New polypeptides and their encoded proteins, useful as nutritional

PT sources or supplements, or in gene therapy, particularly for treating

PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or

PT inflammation -

XX Claim 20; SEQ ID # 759; 357bp + sequence listing; English.

XX The invention relates to 446 newly isolated polynucleotide sequences.

XX The activity of polynucleotides of the invention may be described as,

XX vulnery, neuroprotective, immunomodulator, cytoskeletal and

XX anti-inflammatory. Compositions comprising nucleic acids of the invention

XX are useful for treating a mammalian subject, or as nutritional sources or

XX supplements. These are useful in gene therapy, particularly for treating

XX wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,

CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or  
CC inflammation. The nucleic acids and polypeptides are also useful in  
CC diagnostic and research methods. The sequences given in records in  
CC ABP43544-ABP43589 represent polypeptides encoded by polynucleotides of  
CC the invention.  
CC NOTE: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
XX Sequence 304 AA;

Query Match 71.6%; Score 101; DB 22; Length 304;  
Best Local Similarity 89.5%; Pred. No. 3.9e-05;  
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 6 LKRGCGGGGGGGGGGGG 24  
|:||||||||||||||  
Db 99 LRAGGGGGGGGGGGGGG 117

RESULT 10  
AAM52318  
ID AAM52318 standard; Protein: 520 AA.

XX AAM52318;  
XX  
XX 18-JAN-2002 (first entry)  
XX  
XX

DE Murine WASP protein.

XX Actin polymerisation; Ena/VASP; vasodilator-stimulated phosphoprotein;  
XX metastatic cancer; parasitic infection; cytotoxic; Murine; WASP.

XX Mus musculus.

OS  
XX WO20017356-A2.

PN  
XX 27-SEP-2001.

PD  
XX 21-MAR-2001; 2001WO-FR00843.

PF  
XX 22-MAR-2000; 2000FR-0003637.

PR  
XX (CNRS ) CENT NAT RECH SCI.  
PA (CURT-) INST CURIE.

XX  
XX Fradelizi J, Friederich E, Golsteyn RM, Louvard D, Noireaux V;  
PI Sykes C;

DR WPI; 2001-639148/73.

XX  
XX Identifying modulators of actin polymerization, potentially useful for  
PT treating tumor metastasis and parasitic infection, using proteins that  
PT contain Ena/VASP binding sites -

XX  
XX Claim 13; Pages 100-101; 109pp; French.

XX The present invention relates to a method for identifying modulators of  
CC actin polymerisation. The method involves using proteins that contain at  
CC least one binding motif for proteins of the Ena/VASP  
CC (vasodilator-stimulated phosphoprotein) family in the preparation of  
CC reagents for identification/screening of molecules that modulate  
CC formation of the actin cytoskeleton. The proteins used in the method  
CC (i.e. the proteins with binding motif(s) for Ena/VASP proteins) do not  
CC bind to the Arp2/3 protein complex. The modulators identified by the  
CC method are potentially useful for treating disorders of actin  
CC polymerisation, e.g. metastatic cancer or parasitic infection, and as  
CC cytotoxic agents. The present sequence one such protein with binding  
CC motif(s) for Ena/VASP proteins, which was used in the method of the  
CC present invention.

XX  
XX Sequence 520 AA;

Query Match 71.6%; Score 101; DB 22; Length 520;  
Best Local Similarity 100.0%; Pred. No. 6.3e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 RGGGGGGGGGGGGGGG 24  
|||||||  
Db 323 RGGGGGGGGGGGGGGG 339

RESULT 11  
AAG67350  
ID AAG67350 standard; Protein: 520 AA.

XX AAG67350;  
XX

DT 13-NOV-2001 (first entry)

DE Amino acid sequence of a murine WASP protein.

XX Wiskott-Aldrich syndrome protein; WASP; actin cytoskeleton;  
XX cell motility; actin polymerisation; cancer; parasitic infection;  
XX embryonic development; immune response; wound repair.

XX Mus musculus.

OS  
XX WO200144292-A2.

PN  
XX 21-JUN-2001.

PD  
XX 15-DEC-2000; 2000WO-FR03569.

PF  
XX 16-DEC-1999; 99FR-0015900.

PR  
XX (CNRS ) CENT NAT RECH SCI.  
PA (CURT-) INST CURIE.

XX  
XX Noireaux V, Prost J, Sykes C, Friederich E, Golsteyn RM;  
PI Louvard D;

DR WPI; 2001-536241/59.

DR N-PSDB; AAT77902.

XX  
XX New fragments of WASP family proteins, useful for detecting and  
PT identifying modulators of actin cytoskeleton formation, potential  
PT anticancer and antiparasitic agents -

XX  
XX Claim 10; Fig 4; 162pp; French.

XX The present sequence represents WASP (Wiskott-Aldrich syndrome protein).  
CC Peptide fragments of WASP-family proteins of eukaryotic cells are used  
CC to prepare reagents for detecting compounds that inhibit or stimulate  
CC formation of the actin cytoskeleton, and thus inhibit or stimulate cell  
CC motility. The peptides are used to detect and identify compounds which  
CC are potentially useful for treating diseases associated with dysfunction  
CC of actin polymerisation, particularly metastatic cancer and parasite  
CC infection; as cytotoxic agents for inhibiting/stimulating formation of  
CC the actin cytoskeleton and for detecting side-effects, on actin  
CC polymerisation, of pharmaceuticals. By modulating actin polymerisation,  
CC these compounds affect cell motility, embryonic development, the immune  
CC response and wound repair.

XX  
XX Sequence 520 AA;

Query Match 71.6%; Score 101; DB 22; Length 520;  
Best Local Similarity 100.0%; Pred. No. 6.3e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 RGGGGGGGGGGGGGGG 24  
|||||||  
Db 323 RGGGGGGGGGGGGGGG 339

RESULT 12

AAB49336  
 ID AAB49336 standard; peptide; 520 AA.  
 XX  
 AC AAB49336;  
 XX  
 DT 01-MAR-2001 (first entry)  
 XX  
 DE Murine WASP protein.  
 XX  
 KM Same8; arginine methylation; arginine methyltransferase activity; mouse;  
 KW cancer; Wiskott-Aldrich syndrome; antibody; SLM-1; SLM-2; WASP; hnRNP K.  
 XX  
 OS Mus sp.  
 XX  
 PN CA2266760-A1.  
 XX  
 PD 08-OCT-2000.  
 XX  
 PF 08-APR-1999; 99CA-2266760.  
 XX  
 PR 08-APR-1999; 99CA-2266760.  
 XX  
 PA (DAVI-) DAVIS-JEWISH GEN HOSPITAL MORTIMER B.  
 XX  
 PI Richard S;  
 XX  
 DR WPI; 2001-050180/07.  
 XX  
 PT Assaying arginine methyltransferase activity useful for determining  
 PT whether a cell is cancerous or has cancer potential comprises measuring  
 PS the methylation of specific substrates such as SLM-1, SLM-2, Same8 -  
 XX  
 PS Disclosure; Fig 5; 25pp; English.  
 XX  
 CC The present invention provides a novel method for assaying arginine  
 CC methyltransferase activity in a cell by measuring the methylation of  
 CC certain proteins. These proteins are Same8, SLM-1, SLM-2, hnRNP K and  
 CC WASP. Also provided are antibodies capable of recognising proteins  
 CC containing methylated arginines. These can be used in the stimulation of  
 CC an immune response and to measure the degree of methylation of the  
 CC proteins. The method is useful in the diagnosis of cancer; cancerous  
 CC cells and Wiskott-Aldrich syndrome. It is also useful for measuring the  
 CC extent to which ligands are binding to receptors.  
 CC  
 SQ Sequence 520 AA;  
 XX  
 QY  
 8 RGGGGGGGGGGGGGGG 24  
 |||||  
 323 RGGGGGGGGGGGGGGG 339  
 DB  
 RESULT 13  
 AAY57393  
 ID AAY57393 standard; Protein; 688 AA.  
 XX  
 AC AAY57393;  
 XX  
 DT 23-JUN-2000 (first entry)  
 XX  
 DE Mouse nerve differentiation factor, Tbr-2.  
 XX  
 KM Tbr-2; nerve differentiation factor; brain; nerve disorder; mouse;  
 KW senile dementia; nerve regeneration; nerve injury.  
 XX  
 OS Mus musculus.  
 XX  
 PN MO200012712-A1.  
 XX  
 PD 09-MAR-2000.  
 XX

XX  
 PF 18-JUN-1999; 99MO-JP03271.  
 XX  
 PR 26-AUG-1998; 98JP-0240920.  
 XX  
 PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.  
 XX  
 PI Kimura N, Taga T;  
 XX  
 DR WPI; 2000-237874/20.  
 DR N-PSDB; AA290784.  
 XX  
 PT Gene Tbr-2 highly expressed in fetal brain tissue encodes a nerve  
 PT differentiation factor for treatment of nerve disorders such as  
 PT dementia and nerve injury -  
 XX  
 PS Claim 3; Page 62-70; 89pp; Japanese.  
 XX  
 CC The invention relates to a gene Tbr-2 encoding a nerve differentiation  
 CC factor that is highly expressed in fetal brain tissue. The nerve  
 CC differentiation factor polypeptides can be expressed by standard  
 CC recombinant methodology. The nerve differentiation factor is useful for  
 CC the treatment of nerve disorders such as senile dementia, and for the  
 CC promotion of nerve regeneration after nerve injury. The present sequence  
 CC represents a mouse nerve differentiation factor, Tbr-2.  
 CC  
 SQ Sequence 688 AA;  
 XX  
 QY  
 1 PPVHLKGGGGGGGGGGGG 23  
 |||||  
 19 FYSLSARGGGGGGGGGGGG 41  
 DB  
 RESULT 14  
 AAM13632  
 ID AAM13632 standard; peptide; 50 AA.  
 XX  
 AC AAM13632;  
 XX  
 DT 17-DEC-1997 (first entry)  
 XX  
 DE Apolipoprotein (a) secondary standard exemplary structure.  
 XX  
 KM Apolipoprotein; immunogen; antibody; lipoprotein; immunoassay; human.  
 XX  
 OS Synthetic.  
 XX  
 FH Key  
 FT 1..19  
 FT /label= X  
 FT /note= "Protease cleavage region of apolipoprotein (a)"  
 FT Modified-site  
 FT 20..39  
 FT /label= Z  
 FT /note= "Acetylated; written as Aco-Val-"  
 FT Region  
 FT 40..50  
 FT /label= Y  
 FT /note= "C-terminal end of apolipoprotein B-100"  
 FT  
 XX  
 PN EP764657-A1.  
 XX  
 PD 26-MAR-1997.  
 XX  
 PF 19-SEP-1995; 95EP-0306592.  
 XX  
 PR 19-SEP-1995; 95EP-0306592.  
 XX  
 PA (CARB/) CARBAUGH J E.  
 XX



```
XX
PI Chiknaas SG;
XX
DR WPI; 1997-181788/17.
XX
PT Apo:lipo:protein (a) peptide(s) and their immunogenic conjugates -
PT for produ. of antibodies to apo:lipo:protein (a)
XX
PS Example 11; Page 16; 24pp; English.
XX
CC The present sequence represents an exemplary structure for an
CC apolipoprotein (a) secondary standard. Antibodies which bind to the
CC apolipoprotein (a) activation site epitope are useful as immunoassay
CC reagents for the determination of lipoprotein (a) or apolipoprotein (a).
CC The antibodies or their conjugates, e.g. with agents that cleave
CC lipoprotein (a), can be used in therapeutic compositions for lowering
CC lipoprotein (a) levels. The peptide and construct are useful for
CC provoking an immune response to apolipoprotein (a) in mammals.
XX
SQ Sequence 50 AA;
XX
Query Match 70.2%; Score 99; DB 18; Length 50;
Best Local Similarity 85.7%; Pred.No. 1.4e-05;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
QY 4 VHLKRGGGGGGGGGGGGGGGG 24
Db 15 VALTPGGGGGGGGGGGGGGG 35
XX
RESULT 15
AAM13633
ID AAM13633 standard; peptide; 50 AA.
XX
AC AAM13633;
XX
DT 17-DEC-1997 (first entry)
XX
DE Apolipoprotein (a) secondary standard exemplary structure.
XX
KM Apolipoprotein; immunogen; antibody; lipoprotein; immunoassay; human.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Region 1..19 /label=X
FT Modified-site 1 /note="Protease cleavage region of apolipoprotein (a)"
FT Region 20..39 /note="Acetylated; written as ACO-Val-"
FT FT /label=Z
FT FT /note="Spacer region to separate the epitopes"
FT Region 40..50 /label=Y
FT FT /note="C-terminal end of apolipoprotein (a)"
XX
EP764657-A1.
XX
PD 26-MAR-1997.
XX
PF 19-SEP-1995; 95EP-0306592.
XX
PR 19-SEP-1995; 95EP-0306592.
XX
PA (CARB/) CARBAUGH J E.
XX
PI Chiknaas SG;
XX
DR WPI; 1997-181788/17.
XX
PT Apo:lipo:protein (a) peptide(s) and their immunogenic conjugates -
PT for produ. of antibodies to apo:lipo:protein (a)
```

```
XX
PS Example 11; Page 17; 24pp; English.
XX
CC The present sequence represents an exemplary structure for an
CC apolipoprotein (a) secondary standard. Antibodies which bind to the
CC apolipoprotein (a) activation site epitope are useful as immunoassay
CC reagents for the determination of lipoprotein (a) or apolipoprotein (a).
CC The antibodies or their conjugates, e.g. with agents that cleave
CC lipoprotein (a), can be used in therapeutic compositions for lowering
CC lipoprotein (a) levels. The peptide and construct are useful for
CC provoking an immune response to apolipoprotein (a) in mammals.
XX
SQ Sequence 50 AA;
XX
Query Match 70.2%; Score 99; DB 18; Length 50;
Best Local Similarity 85.7%; Pred.No. 1.4e-05;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
QY 4 VHLKRGGGGGGGGGGGGGGGG 24
Db 15 VALTPGGGGGGGGGGGGGGG 35
XX
Search completed: December 3, 2003, 15:49:55
Job time : 23.2041 sec
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 15:47:35 ; Search time 7.67247 Seconds  
(without alignments)  
132.334 Million cell updates/sec

Title: US-09-788-308D-4

Perfect score: 141

Sequence: 1 FFPVHLKRGCGGCGGCGGCGG 24

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	98	69.5	33 4 US-09-039-780A-24	Sequence 24, Appl
2	96	68.1	20 1 US-08-234-602-14	Sequence 14, Appl
3	96	68.1	20 4 US-09-039-780A-7	Sequence 7, Appl
4	96	68.1	161 2 US-08-581-528A-6	Sequence 6, Appl
5	96	68.1	161 5 PCT-US94-07799-6	Sequence 6, Appl
6	96	68.1	201 3 US-09-052-995-1	Sequence 1, Appl
7	96	68.1	201 3 US-09-053-003-40	Sequence 40, Appl
8	96	68.1	201 4 US-09-054-281-22	Sequence 22, Appl
9	96	68.1	201 4 US-09-478-948-6	Sequence 6, Appl
10	96	68.1	240 1 US-08-362-670B-30	Sequence 30, Appl
11	96	68.1	240 3 US-08-333-576C-30	Sequence 30, Appl
12	96	68.1	240 3 US-08-808-324-30	Sequence 30, Appl
13	96	68.1	240 5 PCT-US94-14030A-30	Sequence 30, Appl
14	96	68.1	266 3 US-09-032-523-7	Sequence 7, Appl
15	96	68.1	268 2 US-08-835-099A-9	Sequence 9, Appl
16	96	68.1	268 2 US-09-157-349-9	Sequence 9, Appl
17	96	68.1	918 3 US-09-041-886-11	Sequence 11, Appl
18	96	68.1	1084 4 US-09-394-272-9	Sequence 9, Appl
19	90	63.8	148 1 US-08-207-904-15	Sequence 15, Appl
20	90	63.8	322 1 US-08-014-943A-2	Sequence 2, Appl
21	90	63.8	322 1 US-08-486-421-3	Sequence 3, Appl
22	90	63.8	322 1 US-08-470-911-3	Sequence 3, Appl
23	90	63.8	322 2 US-08-486-809-3	Sequence 3, Appl
24	90	63.8	738 3 US-08-864-038A-3	Sequence 3, Appl
25	88.5	62.8	501 4 US-09-342-325C-44	Sequence 44, Appl
26	87	61.7	354 4 US-09-393-627B-28	Sequence 28, Appl
27	86	61.0	1349 3 US-08-938-291A-6	Sequence 6, Appl

28	86	61.0	1349 4 US-09-589-619-6	Sequence 6, Appl
29	85	60.3	28 2 US-08-400-115-21	Sequence 21, Appl
30	85	60.3	239 2 US-08-860-174A-2	Sequence 2, Appl
31	85	60.3	241 3 US-08-902-486-13	Sequence 13, Appl
32	85	60.3	241 4 US-09-554-765-13	Sequence 13, Appl
33	85	60.3	258 2 US-08-860-174A-13	Sequence 13, Appl
34	85	60.3	271 2 US-08-400-115-4	Sequence 4, Appl
35	85	60.3	408 4 US-09-554-765-15	Sequence 15, Appl
36	85	60.3	409 4 US-09-554-765-14	Sequence 14, Appl
37	85	60.3	496 3 US-08-902-486-15	Sequence 15, Appl
38	85	60.3	666 4 US-09-423-439-51	Sequence 51, Appl
39	85	60.3	673 4 US-09-423-439-32	Sequence 32, Appl
40	85	60.3	1958 1 US-07-945-283-2	Sequence 2, Appl
41	84	59.6	22 4 US-08-406-824A-24	Sequence 24, Appl
42	84	59.6	211 2 US-08-935-886-10	Sequence 10, Appl
43	84	59.6	360 3 US-08-319-866-2	Sequence 2, Appl
44	84	59.6	400 2 US-09-086-010-2	Sequence 2, Appl
45	83.5	59.2	342 4 US-09-252-991A-19903	Sequence 19903, A

#### ALIGNMENTS

RESULT 1  
US-09-039-780A-24  
Sequence 24, Application US/09039780A  
Patent No. 6376248  
GENERAL INFORMATION:  
APPLICANT: HAMLEY-NEILSON, PAMELA  
LAN, JIANQING  
SHH, POJEN  
JESSE, JOEL A.  
SCHIFFERLI, KEVIN P.  
GEBERHU, GUILIAT  
TITLE OF INVENTION: PEPTIDE-ENHANCED TRANSFECTIONS  
NUMBER OF SEQUENCES: 120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GREENLEE, WINNER & SULLIVAN  
STREET: 5370 MANHATTAN CIRCLE, SUITE 201  
CITY: BOULDER  
STATE: CO  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/039,780A  
FILING DATE: 16-Mar-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: SULLIVAN, SALLY A.  
REGISTRATION NUMBER: 32,064  
REFERENCE/DOCKET NUMBER: 32-95C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /product= "OTHER"  
/note= "C AT POSITION 1 CAN BE ABSENT"  
FEATURE:  
NAME/KEY: Modified-site

LOCATION: 3  
OTHER INFORMATION: /product= "OTHER"  
/note= "Xaa AT POSITION 3 CAN BE TYR OR TRP OR CAN BE ABSENT"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 14.33  
OTHER INFORMATION: /product= "OTHER"  
/note= "ANY OR ALL OF THE GLY AT POSITIONS 14 TO 33 CAN BE ABSENT"  
SEQUENCE DESCRIPTION: SEQ ID NO: 24:  
US-09-039-780A-24

Query Match 69.5%; Score 98; DB 4; Length 33;  
Best Local Similarity 94.4%; Pred. No. 2.5e-06;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 KRGGGGGGGGGGGGGG 24  
10 KVGGGGGGGGGGGGGG 27

RESULT 2  
US-08-234-602-14  
Sequence 14, Application US/08234602  
Patent No. 5490981  
GENERAL INFORMATION:  
APPLICANT: CHIKNAS, Steven G.  
TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS  
TITLE OF INVENTION: AND METHODS FOR LIPOPROTEIN(a)  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
City: Alexandria  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/234,602  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/086,358  
FILING DATE: 06-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/832,994  
FILING DATE: 10-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: SANDERCOCK, Colin G.  
REGISTRATION NUMBER: 31,298  
REFERENCE/DOCKET NUMBER: 32879/116/CAJO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-234-602-14

Query Match 68.1%; Score 96; DB 1; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.8e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 GGGGGGGGGGGGGG 24

Db 1 GGGGGGGGGGGGGG 16

RESULT 3  
US-09-039-780A-7  
Sequence 7, Application US/09039780A  
Patent No. 6376248  
GENERAL INFORMATION:  
APPLICANT: HAWLEY-NELSON, PAMELA  
LAN, JIANQING  
SHIH, POJEN  
JESSE, JOEL A.  
SCHEFFERLI, KEVIN P.  
GEBBYERU, GUILIAT  
TITLE OF INVENTION: PEPTIDE-ENHANCED TRANSFECTIONS  
NUMBER OF SEQUENCES: 120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GREENLEE, WINNER & SULLIVAN  
STREET: 5370 MANHATTAN CIRCLE, SUITE 201  
CITY: BOULDER  
STATE: CO  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/039,780A  
FILING DATE: 16-Mar-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: SULLIVAN, SALLY A.  
REGISTRATION NUMBER: 32,064  
REFERENCE/DOCKET NUMBER: 32-95C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303)499-8080  
TELEFAX: (303)499-8089

INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 2..20  
OTHER INFORMATION: /product= "OTHER"  
/note= "ANY OR ALL OF THE AMINO ACIDS 2-20 CAN BE PRESENT OR ABSENT"  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
US-09-039-780A-7

Query Match 68.1%; Score 96; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 2.8e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GGGGGGGGGGGGGG 24  
Db 1 GGGGGGGGGGGGGG 16

RESULT 4  
US-08-581-528A-6  
Sequence 6, Application US/08581528A  
Patent No. 5986058  
GENERAL INFORMATION:  
APPLICANT: Lee, Se-jin  
Huyhn, Thanh  
TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-7

```

; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/581,528A
; FILING DATE: 03-Sept-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/089,670
; FILING DATE: 09-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lisa A. Haile, Ph.D.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/081001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-581-528A-6

Query Match      68.1%; Score 96; DB 2; Length 161;
Best Local Similarity 100.0%; Pred. No. 2e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 GGGGGGGGGGGGGGGG 24
DB      28 GGGGGGGGGGGGGGGG 43

RESULT 5
PCT-US94-07799-6
; Sequence 6, Application PC/TUS9407799
; GENERAL INFORMATION:
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE
; TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-7
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spensley Horn Jubas & Lubitz
; STREET: 1880 Century Park East, Suite 500
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07799
; FILING DATE: 08-JUL-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: TUMARKIN, LISA A., PH.D.
; REGISTRATION NUMBER: P-38,347
; REFERENCE/DOCKET NUMBER: FD-2348
; TELECOMMUNICATION INFORMATION:
```

```

; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-07799-6

Query Match      68.1%; Score 96; DB 5; Length 161;
Best Local Similarity 100.0%; Pred. No. 2e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 GGGGGGGGGGGGGGGG 24
DB      28 GGGGGGGGGGGGGGGG 43

RESULT 6
US-09-052-995-1
; Sequence 1, Application US/09052995
; Patent No. 6183956
; GENERAL INFORMATION:
; APPLICANT: Sivaraia, Mohanram
; APPLICANT: Strulovici, Berta
; APPLICANT: Flores, Osvaldo A.
; TITLE OF INVENTION: High Throughput In Vitro Screening Assay
; TITLE OF INVENTION: for Transcription Modulators
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/052,995
; FILING DATE: 31-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parent, Annette S.
; REGISTRATION NUMBER: 42,058
; REFERENCE/DOCKET NUMBER: 018781-000600US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-052-995-1

Query Match      68.1%; Score 96; DB 3; Length 201;
Best Local Similarity 100.0%; Pred. No. 2.5e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 GGGGGGGGGGGGGGGG 24
DB      1 GGGGGGGGGGGGGGGG 16

RESULT 7
US-09-053-003-40
```

Sequence 40, Application US/09053003  
Patent No. 6207391  
GENERAL INFORMATION:  
APPLICANT: Wu, Pengguang  
APPLICANT: McKinney, Judi  
TITLE OF INVENTION: High-Throughput Screening Assays for  
TITLE OF INVENTION: Modulators of STAT4 and STAT6 Activity  
NUMBER OF SEQUENCES: 64  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/053,003  
FILING DATE: 31-MAR-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, Timothy L.  
REGISTRATION NUMBER: 35,367  
REFERENCE/DOCKET NUMBER: 018781-0008000US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 40:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 201 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
MOLECULE TYPE: linear  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1..97  
OTHER INFORMATION: /product= "OTHER"  
OTHER INFORMATION: /note= "Gly at positions 1-97 may be  
OTHER INFORMATION: present or absent"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 105..201  
OTHER INFORMATION: /product= "OTHER"  
OTHER INFORMATION: /note= "Gly at positions 105-201 may be  
OTHER INFORMATION: present or absent"  
US-09-053-003-40  
Query Match 68.1%; Score 96; DB 3; Length 201;  
Best Local Similarity 100.0%; Pred. No. 2.5e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 GGGGGGGGGGGGGGGG 24  
DB 1 GGGGGGGGGGGGGGGG 16  
RESULT 8  
US-09-054-281-22  
Sequence 22, Application US/09054281  
Patent No. 6444421  
GENERAL INFORMATION:  
APPLICANT: Chung, Jay H.  
TITLE OF INVENTION: Methods for Detecting Intermolecular  
TITLE OF INVENTION: Interactions in Vivo and in Vitro  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/054,281  
FILING DATE: 02-APR-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/080,234  
FILING DATE: 03-APR-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/826,622  
FILING DATE: 03-APR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Smith, Timothy L.  
REGISTRATION NUMBER: 35,367  
REFERENCE/DOCKET NUMBER: 015280-2952000US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 201 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
MOLECULE TYPE: linear  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1..97  
OTHER INFORMATION: /product= "OTHER"  
OTHER INFORMATION: /note= "Gly at positions 1-97 may be  
OTHER INFORMATION: present or absent"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 105..201  
OTHER INFORMATION: /product= "OTHER"  
OTHER INFORMATION: /note= "Gly at positions 105-201 may be  
OTHER INFORMATION: present or absent"  
US-09-054-281-22  
Query Match 68.1%; Score 96; DB 4; Length 201;  
Best Local Similarity 100.0%; Pred. No. 2.5e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 GGGGGGGGGGGGGGGG 24  
DB 1 GGGGGGGGGGGGGGGG 16  
RESULT 9  
US-09-478-948-6  
Sequence 6, Application US/09478948  
Patent No. 6465258  
GENERAL INFORMATION:  
APPLICANT: Shan, Bei  
APPLICANT: Okamoto, Arthur Y.  
APPLICANT: Tularik Inc.  
TITLE OF INVENTION: FXR Receptor-Mediated Modulation of Cholesterol  
TITLE OF INVENTION: Metabolism  
FILE REFERENCE: 018781-0013100US  
CURRENT APPLICATION NUMBER: US/09/478,948  
CURRENT FILING DATE: 2000-01-06  
PRIOR APPLICATION NUMBER: US 60/115,249  
PRIOR FILING DATE: 1999-01-07  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 6  
LENGTH: 201  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: flexible linker  
NAME/KEY: MOD\_RES  
LOCATION: (1)-(97)  
OTHER INFORMATION: Gly at positions 1-97 may be present or absent  
NAME/KEY: MOD\_RES  
LOCATION: (105)-(201)  
OTHER INFORMATION: Gly at positions 105-201 may be present or absent  
US-09-478-948-6

Query Match  
Best Local Similarity 100.0%; Score 96; DB 4; Length 201;  
Pred. No. 2.5e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GGGGGGGGGGGGGGGG 24  
DB 1 GGGGGGGGGGGGGGGG 16

RESULT 10  
US-08-362-670B-30  
Sequence 30, Application US/08362670B  
Patent No. 5658882  
GENERAL INFORMATION:  
APPLICANT: Celeste, Anthony J.  
APPLICANT: Wozney, John  
APPLICANT: Rosen, Vicki A.  
APPLICANT: Wolfman, Neil  
APPLICANT: Thomsen, Gerald H.  
APPLICANT: Mellon, Douglas A.  
TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENETICS INSTITUTE, INC.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/362,670B  
FILING DATE: December 22, 1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Lazat, Steven R.  
REGISTRATION NUMBER: 32,618  
REFERENCE/DOCKET NUMBER: 5202-D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617 498-8260  
TELEFAX: 617 876-5851  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 240 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-362-670B-30

Query Match  
Best Local Similarity 100.0%; Score 96; DB 1; Length 240;  
Pred. No. 3e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GGGGGGGGGGGGGGGG 24  
DB 1 GGGGGGGGGGGGGGGG 16

DB 107 GGGGGGGGGGGGGGGG 122

RESULT 11  
US-08-333-576C-30  
Sequence 30, Application US/08333576C  
Patent No. 6027919  
GENERAL INFORMATION:  
APPLICANT: Celeste, Anthony J.  
APPLICANT: Wozney, John  
APPLICANT: Rosen, Vicki A.  
APPLICANT: Wolfman, Neil  
APPLICANT: Thomsen, Gerald H.  
APPLICANT: Mellon, Douglas A.  
TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENETICS INSTITUTE, INC.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/333,576C  
FILING DATE: No. 6027919eember 2, 1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lazat, Steven R.  
REGISTRATION NUMBER: 32,618  
REFERENCE/DOCKET NUMBER: 5202-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617 498-8260  
TELEFAX: 617 876-5851  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 240 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-333-576C-30

Query Match  
Best Local Similarity 100.0%; Score 96; DB 3; Length 240;  
Pred. No. 3e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GGGGGGGGGGGGGGGG 24  
DB 107 GGGGGGGGGGGGGGGG 122

RESULT 12  
US-08-808-324-30  
Sequence 30, Application US/08808324  
Patent No. 6284872  
GENERAL INFORMATION:  
APPLICANT: Celeste, Anthony J.  
APPLICANT: Wozney, John  
APPLICANT: Rosen, Vicki A.  
APPLICANT: Wolfman, Neil  
APPLICANT: Thomsen, Gerald H.  
APPLICANT: Mellon, Douglas A.  
TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENETICS INSTITUTE, INC.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge

STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/808,324  
FILING DATE: Herewith  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Lazari, Steven R.  
REGISTRATION NUMBER: 32,618  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617 498-8260  
TELEFAX: 617 876-5851  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 240 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-808-324-30

Query Match 68.1%; Score 96; DB 3; Length 240;  
Best Local Similarity 100.0%; Pred. No. 3e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GGGGGGGGGGGGGG 24  
Db 107 GGGGGGGGGGGGGG 122

RESULT 13  
PCT-US94-14030A-30  
Sequence 30, Application PC/TUS9414030A  
GENERAL INFORMATION:  
APPLICANT: GENETICS INSTITUTE, INC.  
TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENETICS INSTITUTE, INC.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/14030A  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/164,103  
FILING DATE: 07-DEC-1993  
APPLICATION NUMBER: US 08/217,780  
FILING DATE: 25-MAR-1994  
APPLICATION NUMBER: US 08/333,576  
FILING DATE: 02-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lazari, Steven R.  
REGISTRATION NUMBER: 32,618  
REFERENCE/DOCKET NUMBER: 5202D-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617 498-8260

TELEFAX: 617 876-5851  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 240 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-14030A-30

Query Match 68.1%; Score 96; DB 5; Length 240;  
Best Local Similarity 100.0%; Pred. No. 3e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GGGGGGGGGGGGGG 24  
Db 107 GGGGGGGGGGGGGG 122

RESULT 14  
US-09-032-523-7  
Sequence 7, Application US/09032523  
Patent No. 6232454  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
APPLICANT: Guegler, Karl  
APPLICANT: Baugh, Mariah  
TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/032,523  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0479 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 266 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 164403  
US-09-032-523-7

Query Match 68.1%; Score 96; DB 3; Length 266;  
Best Local Similarity 100.0%; Pred. No. 3e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GGGGGGGGGGGGGG 24

Db 37 GGGGGGGGGGGGGGGG 52

RESULT 15

US-08-835-099A-9  
Sequence 9, Application US/08835099A

Patent No. 5874277

GENERAL INFORMATION:

APPLICANT: SHINTANI, Yasushi

APPLICANT: KAWAMOTO, Tomohiro

TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION

TITLE OF INVENTION: AND USE

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/835,099A

FILING DATE: 04-APR-1997

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 083649/1996

FILING DATE: 05-APR-1996

APPLICATION NUMBER: 92105508.2

FILING DATE: 03-APR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Resnick, David S

REGISTRATION NUMBER: 34,235

REFERENCE/DOCKET NUMBER: 47342

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

TELEX:

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 268 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-835-099A-9

Query Match 68.1%; Score 96; DB 2; Length 268;

Best Local Similarity 100.0%; Pred. No. 3.3e-05;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 GGGGGGGGGGGGGGGG 24

Db 37 GGGGGGGGGGGGGGGG 52

Search completed: December 3, 2003, 15:54:26

Job time : 7.67347 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 15:50:00 ; Search time 14.5306 Seconds  
(without alignments)  
307.186 Million cell updates/sec

Title: US-09-788-308D-4

Perfect score: 141  
Sequence: 1 FFPVHLKRGCGGGCGGGCGGG 24

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Published Applications AA.\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	141	100.0	24	11	US-09-788-308D-4
2	101	71.6	100	9	US-09-893-600-3
3	101	71.6	100	9	US-09-893-499-3
4	101	71.6	100	10	US-09-894-030-3
5	101	71.6	243	12	US-10-029-386-32275
6	101	71.6	520	12	US-10-168-097A-36
7	101	71.6	520	12	US-10-239-431A-34
8	99	70.2	282	15	US-10-197-666A-48
9	98	69.5	33	11	US-09-911-569-24
10	98	69.5	33	12	US-10-200-879-24
11	98	69.5	2783	10	US-09-816-669A-14
12	97	68.8	154	15	US-10-178-213-374
13	96	68.1	20	11	US-09-911-569-7
14	96	68.1	20	12	US-10-200-879-7
15	96	68.1	101	9	US-09-864-761-33599

16	96	68.1	151	12	US-10-164-279-61
17	96	68.1	200	10	US-09-798-584-18
18	96	68.1	200	10	US-09-967-624-19
19	96	68.1	200	10	US-09-998-667-18
20	96	68.1	200	11	US-09-990-940-21
21	96	68.1	200	11	US-09-989-981A-13
22	96	68.1	200	11	US-09-850-948-29
23	96	68.1	200	12	US-10-100-818-14
24	96	68.1	200	12	US-10-293-582-27
25	96	68.1	200	12	US-10-237-467-18
26	96	68.1	200	12	US-10-321-204-54
27	96	68.1	200	12	US-10-179-766-12
28	96	68.1	200	12	US-10-123-568-4
29	96	68.1	200	12	US-10-123-568-4
30	96	68.1	200	12	US-10-135-724-5
31	96	68.1	200	15	US-10-160-354-4
32	96	68.1	200	15	US-10-026-021-8
33	96	68.1	200	15	US-10-161-165-3
34	96	68.1	200	15	US-10-160-663-3
35	96	68.1	200	15	US-10-071-838-15
36	96	68.1	200	15	US-10-094-417-25
37	96	68.1	200	15	US-10-188-405-13
38	96	68.1	200	15	US-10-273-575-29
39	96	68.1	200	15	US-10-233-098-5
40	96	68.1	200	15	US-10-245-850-3
41	96	68.1	201	9	US-09-848-990-22
42	96	68.1	201	10	US-09-760-364-14
43	96	68.1	240	10	US-09-945-182-30
44	96	68.1	280	12	US-10-259-165-212
45	96	68.1	646	10	US-09-964-899-53

## ALIGNMENTS

RESULT 1  
US-09-788-308D-4  
; Sequence 4, Application US/09788308D  
; Publication No. US20030040468A1  
; GENERAL INFORMATION:  
; APPLICANT: No. US20030040468A1theWestern University  
; TITLE OF INVENTION: Polypeptoid Pulmonary Surfactants  
; FILE REFERENCE: 6374  
; CURRENT APPLICATION NUMBER: US/09/788,308D  
; PRIOR FILING DATE: 2001-02-16  
; PRIOR APPLICATION NUMBER: US 60/182,847  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 24  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (9) - (24)  
; OTHER INFORMATION: Fifteen N-substituted glycine residues, each such residue 2-methyl  
; OTHER INFORMATION: Ipropyl substituted.  
US-09-788-308D-4

Query Match 100.0%; Score 141; DB 11; Length 24;  
Best Local Similarity 100.0%; Pred. No. 5.2e-10;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FFPVHLKRGCGGGCGGGCGGG 24  
DB 1 FFPVHLKRGCGGGCGGGCGGG 24

RESULT 2  
US-09-893-600-3  
; Sequence 3, Application US/09893600  
; Patent No. US20020019518A1

```
; GENERAL INFORMATION:
; APPLICANT: Hansen, J. No. US20020019518A1man
; TITLE OF INVENTION: Construction of a Strain of Bacillus subtilis 168 that Displays
; FILE REFERENCE: 108172-00057
; CURRENT APPLICATION NUMBER: US/09/893,600
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/215,449
; PRIOR FILING DATE: 2000-06-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: The peptide sequence of sunA-PG20-SL.
US-09-893-600-3
```

```
Query Match 71.6%; Score 101; DB 9; Length 100;
Best Local Similarity 100.0%; Pred. No. 6.8e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 8 RCGGGGGGGGGGGGGGGG 24
Db 56 RCGGGGGGGGGGGGGGGG 72
```

```
RESULT 3
US-09-893-499-3
; Sequence 3, Application US/09893499
; Patent No. US20020052005A1
; GENERAL INFORMATION:
; APPLICANT: Hansen, J. No. US20020052005A1man
; TITLE OF INVENTION: Construction and Screening of Antibody Display Libraries
; FILE REFERENCE: 108172-00056
; CURRENT APPLICATION NUMBER: US/09/893,499
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/215,449
; PRIOR FILING DATE: 2000-06-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: The peptide sequence of sunA-PG20-SL.
US-09-893-499-3
```

```
Query Match 71.6%; Score 101; DB 9; Length 100;
Best Local Similarity 100.0%; Pred. No. 6.8e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 8 RCGGGGGGGGGGGGGGGG 24
Db 56 RCGGGGGGGGGGGGGGGG 72
```

```
RESULT 4
US-09-894-030-3
; Sequence 3, Application US/09894030
; Patent No. US20020165139A1
; GENERAL INFORMATION:
; APPLICANT: Hansen, J. No. US20020165139A1man
; TITLE OF INVENTION: Construction of a Structural Variant of Sublancin to Facilitate
; FILE REFERENCE: 108172-00058
; CURRENT APPLICATION NUMBER: US/09/894,030
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/215,449
; PRIOR FILING DATE: 2000-06-29
```

```
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: The peptide sequence for sublancin-His Tag.
US-09-894-030-3
```

```
Query Match 71.6%; Score 101; DB 10; Length 100;
Best Local Similarity 100.0%; Pred. No. 6.8e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 8 RCGGGGGGGGGGGGGGGG 24
Db 56 RCGGGGGGGGGGGGGGGG 72
```

```
RESULT 5
US-10-029-386-32275
; Sequence 32275, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32275
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009336.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
; OTHER INFORMATION: SWISSPROT HIT: Q03828, EVALUATE 2.00e-74
US-10-029-386-32275
```

```
Query Match 71.6%; Score 101; DB 12; Length 243;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 8 RCGGGGGGGGGGGGGGGG 24
Db 179 RCGGGGGGGGGGGGGGGG 195
```

```
RESULT 6
US-10-168-097A-36
; Sequence 36, Application US/10168097A
; Publication No. US20030166245A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL OF THE RECHERCHE SCIENTIFIQUE
; APPLICANT: INSTITUT CURIE
; TITLE OF INVENTION: WASP FAMILY PROTEIN FRAGMENTS, AND USES THEREOF
; FILE REFERENCE: IPB99WASP
; CURRENT APPLICATION NUMBER: US/10/168,097A
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 520
```

TYPE: PRT  
ORGANISM: Mus musculus  
US-10-168-097A-36

Query Match 71.6%: Score 101; DB 12; Length 520;  
Best Local Similarity 100.0%; Pred. No. 0.00028;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 RCGGGGGGGGGGGGGG 24  
DB 323 RCGGGGGGGGGGGGGG 339

RESULT 7  
US-10-239-431A-34  
Sequence 34, Application US/10239431A  
Publication No. US20030170726A1  
GENERAL INFORMATION:

APPLICANT: FRADELIZE, JULIE  
APPLICANT: FRIEDERICH, EVELYNE  
APPLICANT: GOLSTEVN, ROY M.  
APPLICANT: LOUVARD, DANIEL  
APPLICANT: NOIREAUX, VINCENT  
APPLICANT: SYKES, CECILE  
TITLE OF INVENTION: PEPTIDE SEQUENCES COMPRISING ONE OR MORE UNITS BINDING  
TO PROTEINS OF THE Ena/VASP FAMILY, AND THEIR USES  
FILE REFERENCE: 0508-1032  
CURRENT APPLICATION NUMBER: US/10/239,431A  
CURRENT FILING DATE: 2002-09-23  
PRIOR APPLICATION NUMBER: PCT/FR01/00843  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: FR 00/03637  
PRIOR FILING DATE: 2000-03-22  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 34  
LENGTH: 520  
TYPE: PRT  
ORGANISM: Murine sp.  
US-10-239-431A-34

Query Match 71.6%: Score 101; DB 12; Length 520;  
Best Local Similarity 100.0%; Pred. No. 0.00028;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 RCGGGGGGGGGGGGGG 24  
DB 323 RCGGGGGGGGGGGGGG 339

RESULT 8  
US-10-197-666A-48

Sequence 48, Application US/10197666A  
Publication No. US20030092037A1  
GENERAL INFORMATION:

APPLICANT: ASAMI KASEI KABUSIKI KAISYA  
TITLE OF INVENTION: ELKI phosphorylation related gene  
FILE REFERENCE: PH-1548US  
CURRENT APPLICATION NUMBER: US/10/197,666A  
CURRENT FILING DATE: 2002-11-18  
PRIOR APPLICATION NUMBER: JP 2001-218204  
PRIOR FILING DATE: 2001-07-18  
PRIOR APPLICATION NUMBER: JP 2001-263450  
PRIOR FILING DATE: 2001-08-31  
PRIOR APPLICATION NUMBER: JP 2002-012176  
PRIOR FILING DATE: 2002-01-21  
PRIOR APPLICATION NUMBER: US 60/305,884  
PRIOR FILING DATE: 2001-07-18  
PRIOR APPLICATION NUMBER: US 60/316,304  
PRIOR FILING DATE: 2001-09-04  
PRIOR APPLICATION NUMBER: US 60/350,027  
PRIOR FILING DATE: 2002-01-23  
NUMBER OF SEQ ID NOS: 156

SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 48  
LENGTH: 282  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-197-666A-48

Query Match 70.2%: Score 99; DB 15; Length 282;  
Best Local Similarity 81.8%; Pred. No. 0.00028;  
Matches 18; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY 3 PVHLKRGGGGGGGGGGGG 24  
DB 212 PTH--SGGGGGGGGGGGGGG 231

RESULT 9  
US-09-911-569-24  
Sequence 24, Application US/09911569  
Publication No. US20030069173A1  
GENERAL INFORMATION:

APPLICANT: HAWLEY-NELSON, PAMELA  
LAN, JIANGJING  
SHIH, POJEN  
JESSE, JOEL A.  
SCHIFFERLI, KEVIN P.  
GEBREYEHU, GULILAT  
TITLE OF INVENTION: PEPTIDE-ENHANCED TRANSFECTIONS  
NUMBER OF SEQUENCES: 120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GREENLEE, WINNER & SULLIVAN  
STREET: 5370 MANHATTAN CIRCLE, SUITE 201  
CITY: BOULDER  
STATE: CO  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/911,569  
FILING DATE: 23-Jul-2001  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/039,780  
FILING DATE: 16-MAR-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: SULLIVAN, SALLY A.  
REGISTRATION NUMBER: 32,064  
REFERENCE/DOCKET NUMBER: 32-95D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303)499-8080  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
MOLECULE TYPE: linear  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /product= "OTHER"  
/note= "C AT POSITION 1 CAN BE ABSENT"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3  
OTHER INFORMATION: /product= "OTHER"  
/note= "Xaa AT POSITION 3 CAN BE TYR OR TRP OR CAN BE ABSENT"

NAME/KEY: Modified-site  
LOCATION: 14...33  
OTHER INFORMATION: /product= "OTHER"  
/note= "ANY OR ALL OF THE GLY AT POSITIONS 14 TO 33 CAN BE ABSENT"  
SEQUENCE DESCRIPTION: SEQ ID NO: 24  
US-09-911-569-24

Query Match 69.5%; Score 98; DB 11; Length 33;  
Best Local Similarity 94.4%; Pred. No. 5.9e-05;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 7 KRGGGGGGGGGGGGGGG 24  
10 KVGGGGGGGGGGGGGGG 27

Db

RESULT 10  
US-10-200-879-24  
Sequence 24, Application US/10200879  
Publication No. US20030144230A1  
GENERAL INFORMATION:  
APPLICANT: HANLEY-NELSON, PAMELA  
LAN, JIANQING  
SHIH, POJEN  
JESSE, JOEL A.  
SCHIFFERLI, KEVIN P.  
GEBREYEHU, GUILIAT  
TITLE OF INVENTION: PEPTIDE-ENHANCED TRANSECTIONS  
NUMBER OF SEQUENCES: 120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GREENLEE, WINNER & SULLIVAN  
STREET: 5370 MANHATTAN CIRCLE, SUITE 201  
CITY: BOULDER  
STATE: CO  
COUNTRY: US  
ZIP: 80303

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentm Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/200.879  
FILING DATE: 23-Jul-2002  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/911,569  
FILING DATE: 23-JUL-2001  
APPLICATION NUMBER: US 09/039,780  
FILING DATE: 16-MAR-1998  
APPLICATION NUMBER: US 08/818,200  
FILING DATE: 14-MAR-1997  
APPLICATION NUMBER: US 08/658,130  
FILING DATE: 04-JUN-1996  
APPLICATION NUMBER: US 08/477,354  
FILING DATE: 07-JUN-1995

ATTORNEY/AGENT INFORMATION:  
NAME: SULLIVAN, SALLY A.  
REGISTRATION NUMBER: 32,064  
REFERENCE/DOCKET NUMBER: 32-95E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303)499-8080  
TELEFAX: (303)499-8089

INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
MOLECULE TYPE: linear  
FEATURE:  
NAME/KEY: Modified-site

LOCATION: 1  
OTHER INFORMATION: /product= "OTHER"  
/note= "C AT POSITION 1 CAN BE ABSENT"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3  
OTHER INFORMATION: /product= "OTHER"  
/note= "Xaa AT POSITION 3 CAN BE TYR OR TRP OR CAN BE ABSENT"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 14...33  
OTHER INFORMATION: /product= "OTHER"  
/note= "ANY OR ALL OF THE GLY AT POSITIONS 14 TO 33 CAN BE ABSENT"  
SEQUENCE DESCRIPTION: SEQ ID NO: 24  
US-10-200-879-24

Query Match 69.5%; Score 98; DB 12; Length 33;  
Best Local Similarity 94.4%; Pred. No. 5.9e-05;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 7 KRGGGGGGGGGGGGGGG 24  
10 KVGGGGGGGGGGGGGGG 27

Db

RESULT 11  
US-09-816-669A-14  
Sequence 14, Application US/09816669A  
Patent No. US20020137019A1  
GENERAL INFORMATION:  
APPLICANT: GARABEDIAN, Michael  
APPLICANT: TAMEJA, Samir  
APPLICANT: HITTELMAN, Adam  
APPLICANT: MARKUS, Steven  
TITLE OF INVENTION: METHOD FOR SCREENING TRANSCRIPTIONAL COREGULATORY PROTEINS OF  
TITLE OF INVENTION: TRANSCRIPTION FACTORS, AND ANDROGEN RECEPTOR TRANSCRIPTIONAL C  
FILE REFERENCE: GARABEDIAN-1.1A  
CURRENT APPLICATION NUMBER: US/09/816,669A  
CURRENT FILING DATE: 2001-03-26  
PRIOR APPLICATION NUMBER: 60/225,618  
PRIOR FILING DATE: 2000-08-15  
PRIOR APPLICATION NUMBER: 60/191,768  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: Patentm version 3.1  
SEQ ID NO 14  
LENGTH: 2783  
TYPE: PRT  
ORGANISM: Human  
US-09-816-669A-14

Query Match 69.5%; Score 98; DB 10; Length 2783;  
Best Local Similarity 76.2%; Pred. No. 0.0025;  
Matches 16; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 4 VHLKRGGGGGGGGGGGGG 24  
2580 LHVPTGGGGGGGGGGGGG 2600

Db

RESULT 12  
US-10-178-213-374  
Sequence 374, Application US/10178213  
Publication No. US20030041348A1  
GENERAL INFORMATION:  
APPLICANT: Simmons, Carl R.  
APPLICANT: Navarro Acevedo, Pedro A.  
APPLICANT: Harwell, Leslie  
APPLICANT: Cahoon, Rebecca  
APPLICANT: McCutchen, Billy Fred  
APPLICANT: Lu, Albert

APPLICANT: Herrmann, Rafael  
APPLICANT: Wong, James  
TITLE OF INVENTION: Defensin Polynucleotides and Methods of  
FILE REFERENCE: 35718/246703  
CURRENT APPLICATION NUMBER: US/10/178,213  
CURRENT FILING DATE: 2002-06-21  
PRIOR APPLICATION NUMBER: 60/300,152  
PRIOR FILING DATE: 2001-06-22  
PRIOR APPLICATION NUMBER: 60/300,241  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 469  
SOFTWARE: FacsSD for Windows Version 4.0  
SEQ ID NO 374  
LENGTH: 154  
TYPE: PRT  
ORGANISM: Oryza sativa  
US-10-178-213-374

Query Match 68.8%; Score 97; DB 15; Length 154;  
Best Local Similarity 94.4%; Pred. No. 0.00028;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 7 KGGGGGGGGGGGGGGGGG 24  
DB 117 KRCGGGGGGGGGGGGGGG 134

RESULT 13  
US-09-911-569-7  
Sequence 7, Application US/0911569  
Publication No. US20030069173A1  
GENERAL INFORMATION:  
APPLICANT: HAMLEY-NELSON, PAMELA  
LAN, JIANQING  
SHI, POJEN  
JESSE, JOEL A.  
SCHIFFERLI, KEVIN P.  
GEBREYEHU, GULILAT  
TITLE OF INVENTION: PEPTIDE-ENHANCED TRANSFECTIONS  
NUMBER OF SEQUENCES: 120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GREENLEE, WINNER & SULLIVAN  
STREET: 5370 MANHATTAN CIRCLE, SUITE 201  
CITY: BOULDER  
STATE: CO  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/911,569  
FILING DATE: 23-Jul-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/039,780  
FILING DATE: 16-MAR-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: SULLIVAN, SALLY A.  
REGISTRATION NUMBER: 32,064  
REFERENCE/DOCKET NUMBER: 32-95D  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303)499-8080  
TELEFAX: (303)499-8089  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
FEATURES:  
NAME/KEY: Modified-site  
LOCATION: 2..20  
OTHER INFORMATION: /product="OTHER"  
/note="ANY OR ALL OF THE AMINO ACIDS 2-20 CAN BE PRESENT OR  
ABSENT"  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
US-09-911-569-7

Query Match 68.1%; Score 96; DB 11; Length 20;  
Best Local Similarity 100.0%; Pred. No. 6.6e-05;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 GGGGGGGGGGGGGGGGGG 24  
DB 1 GGGGGGGGGGGGGGGGGG 16

RESULT 14  
US-10-200-879-7  
Sequence 7, Application US/10200879  
Publication No. US2003014230A1  
GENERAL INFORMATION:  
APPLICANT: HAMLEY-NELSON, PAMELA  
LAN, JIANQING  
SHI, POJEN  
JESSE, JOEL A.  
SCHIFFERLI, KEVIN P.  
GEBREYEHU, GULILAT  
TITLE OF INVENTION: PEPTIDE-ENHANCED TRANSFECTIONS  
NUMBER OF SEQUENCES: 120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GREENLEE, WINNER & SULLIVAN  
STREET: 5370 MANHATTAN CIRCLE, SUITE 201  
CITY: BOULDER  
STATE: CO  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/200,879  
FILING DATE: 23-Jul-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/911,569  
FILING DATE: 23-Jul-2001  
APPLICATION NUMBER: US 09/039,780  
FILING DATE: 16-MAR-1998  
APPLICATION NUMBER: US 08/818,200  
FILING DATE: 14-MAR-1997  
APPLICATION NUMBER: US 08/658,130  
FILING DATE: 04-JUN-1996  
APPLICATION NUMBER: US 08/477,354  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: SULLIVAN, SALLY A.  
REGISTRATION NUMBER: 32,064  
REFERENCE/DOCKET NUMBER: 32-95E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303)499-8080  
TELEFAX: (303)499-8089  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

```
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2..20
OTHER INFORMATION: /product= "OTHER"
/note= "ANY OR ALL OF THE AMINO ACIDS 2-20 CAN BE PRESENT OR
ABSENT"
SEQUENCE DESCRIPTION: SEQ ID NO: 7
US-10-200-879-7

Query Match      68.1%; Score 96; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.6e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 GGGGGGGGGGGGGGGG 24
        |||||
Db      1 GGGGGGGGGGGGGGGG 16

RESULT 15
US-09-864-761-33599
; Sequence 33599, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864, 761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263, 6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: AnnuMax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33599
; LENGTH: 101
; TYPE: PRT
```

```
ORGANISM: Homo sapiens
FEATURE:
; OTHER INFORMATION: MAP TO AP000161.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
US-09-864-761-33599

Query Match      68.1%; Score 96; DB 9; Length 101;
Best Local Similarity 100.0%; Pred. No. 0.00026;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 GGGGGGGGGGGGGGGG 24
        |||||
Db      36 GGGGGGGGGGGGGGGG 51
```

Search completed: December 3, 2003, 15:56:08  
Job time: 14.5306 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 15:45:30 ; Search time 7.02041 seconds  
(without alignments)  
328.763 Million cell updates/sec

Title: US-09-788-308D-4

Perfect score: 141  
Sequence: 1 FPPVHLKRGCGGGGGGGGGG 24

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	101	71.6	221 2	T04592 glycine-rich cell
2	101	71.6	333 2	A39065 homeotic protein E
3	100	70.9	443 1	S29334 transcription fact
4	100	70.9	445 1	S31224 transcription fact
5	99	70.2	165 1	KNR2G1 glycine-rich cell
6	99	70.2	396 2	T49109 glycine-rich prote
7	98	69.5	2783 1	A41948 alpha-fetoprotein
8	97	68.8	272 2	T02745 nucleic acid bindi
9	97	68.8	445 1	A49447 transcription fac
10	96	68.1	106 2	P84797 homeothetical prote
11	96	68.1	151 2	S43296 bone morphogenetic
12	96	68.1	207 2	T07381 glycine-rich prote
13	96	68.1	263 2	A34466 calpain (EC 3.4.22
14	96	68.1	266 1	CIRPL calpain (EC 3.4.22
15	96	68.1	268 1	CIRUL calpain (EC 3.4.22
16	96	68.1	290 2	AD1849 homeothetical prote
17	96	68.1	336 1	S16750 chitinase (EC 3.2.
18	96	68.1	384 1	A26099 glycine-rich cell
19	96	68.1	410 2	I38502 gene Brn-3b protei
20	96	68.1	433 2	S20963 homeotic protein H
21	96	68.1	702 1	VCPVAP coat protein VP1 -
22	96	68.1	910 2	A34721 androgen receptor
23	96	68.1	911 2	B34721 androgen receptor
24	96	68.1	919 2	A39248 androgen receptor
25	96	68.1	1084 2	T04103 sucrose-phosphate
26	95	67.4	378 2	S04336 UI snRNP 70K prote
27	93	66.0	136 2	T29282 hypotheical prote
28	93	66.0	199 2	S16063 acp-22 protein - y
29	93	66.0	199 2	S32224 acp-22 protein - y

30	93	66.0	321 2	A38712 fibrillarlin [valid
31	92	65.2	393 2	T20268 hypotheical prote
32	92	65.2	440 2	S71795 transcription fact
33	92	65.2	495 1	S31223 transcription fact
34	91	64.5	206 2	I53066 gene M-twist prote
35	91	64.5	280 2	A42424 chitinase (EC 3.2.
36	91	64.5	404 2	S54729 RNA-binding protei
37	91	64.5	892 2	T27005 hypotheical prote
38	90	63.8	80 2	T10550 hypotheical prote
39	90	63.8	188 2	S49192 GCR 1 protein - fr
40	90	63.8	220 2	A44805 eggshell protein p
41	90	63.8	255 2	B84777 hypotheical prote
42	90	63.8	266 1	CIRBL calpain (EC 3.4.22
43	90	63.8	268 2	S09860 hypotheical prote
44	90	63.8	273 2	T51145 nucleic acid bindi
45	90	63.8	291 1	S31415 glycine-rich prote

## ALIGNMENTS

RESULT 1  
T04592 glycine-rich cell wall structural protein homolog F23E13.120 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 14-May-1999  
C/Accession: T04592  
R/Beyan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Jesse, T.  
Submitted to the Protein Sequence Database, March 1998  
A/Reference number: Z15378  
A/Accession: T04592  
A/Molecule type: DNA  
A/Residues: 1-221 <BEV>  
A/Cross-references: EMBL:AL022141  
A/Experimental source: cultivar Columbia; BAC clone F23E13  
A/Genetics:  
A/Map position: 4  
A/Note: F23E13.120

Query Match 71.6%; Score 101; DB 2; Length 221;  
Best Local Similarity 100.0%; Pred. No. 0.00018;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 RGGGGGGGGGGGGGGG 24  
DB 152 RGGGGGGGGGGGGGGG 168

## RESULT 2

A39065 homeotic protein EVX2 - human (fragment)  
C/Species: Homo sapiens (man)  
C/Date: 20-Mar-1992 #sequence\_revision 20-Mar-1992 #text\_change 17-Oct-1997  
C/Accession: A39065  
R/D'Esposito, M.; Morelli, F.; Acampora, D.; Migliaccio, E.; Simeone, A.; Boncinelli, E.  
Genomics 10, 43-50, 1991  
A/Title: EVX2, a human homeobox gene homologous to the even-skipped segmentation gene, 1  
A/Reference number: A39065; MUID:9157849; PMID:1675198  
A/Accession: A39065  
A/Status: preliminary; not compared with conceptual translation  
A/Molecule type: DNA  
A/Residues: 1-333 <DAB>  
C/Genetics:  
A/Gene: GDB:EVX2  
A/Cross-references: GDB:127528; OMIM:142991  
A/Map position: 2q24.3-2q31  
C/Keywords: unassigned homeobox proteins; homeobox homology  
F/46-102/Domain: homeobox homology <HOX>

Query Match 71.6%; Score 101; DB 2; Length 333;  
Best Local Similarity 100.0%; Pred. No. 0.00025;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 RGGGGGGGGGGGGGGG 24  
 |||||  
 Db 269 RGGGGGGGGGGGGGGG 285

## RESULT 3

transcription factor Brn-2 [validated] - human  
 S29334  
 N/Alternate names: class III POU domain protein brain-2; transcription factor Oct-3  
 N/Contains: transcription factor Brn-2; transcription factor Oct-5a; transcription facto  
 C/Species: Homo sapiens (man)  
 C/Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text\_change 08-Dec-2000  
 C/Accession: S29334; S05043; S30296  
 R/Schreiber, E.; Tobler, A.; Malipiero, U.; Fontana, A.  
 submitted to the EMBL Data Library April 1992  
 A/Description: The human N-Oct 3 cDNA encodes three neuroectodermal cell lineage restrict  
 A/Reference number: S29334  
 A/Accession: S29334  
 A/Molecule type: mRNA  
 A/Residues: 1-443 <SCH>  
 A/Cross-references: EMBL:Z11933; NID:G35084; PIDN:CAA77990.1; PID:G35085  
 A/Experimental source: tissue-type brain  
 R/He, X.; Treacy, M.N.; Simmons, D.M.; Ingraham, H.A.; Swanson, L.W.; Rosenfeld, M.G.  
 Nature 340, 35-42, 1989  
 A/Title: Expression of a large family of POU-domain regulatory genes in mammalian brain  
 A/Reference number: S05042; MUID:89295573; PMID:2739723  
 A/Accession: S05043  
 A/Status: preliminary; not compared with conceptual translation  
 A/Status: preliminary; not compared with conceptual translation  
 A/Molecule type: mRNA  
 A/Residues: 280-350/351-404 <HEX>  
 A/Cross-references: GB:Z11933; NID:G35084  
 R/Schreiber, E.; Tobler, A.; Malipiero, U.; Schaffner, W.; Fontana, A.  
 Nucleic Acids Res. 21, 253-258, 1993  
 A/Title: cDNA cloning of human N-Oct 3, a nervous-system specific POU domain transcript  
 A/Reference number: S30296; MUID:93181199; PMID:8441633  
 A/Accession: S30296  
 A/Status: nucleic acid sequence not shown  
 A/Residues: 1-25, 'G' 27-443 <SCW>  
 A/Molecule type: mRNA  
 A/Cross-references: EMBL:Z11933  
 A/Experimental source: tissue-type brain  
 A/Genetics:  
 A/Gene: GDB:POU3F2; OTF7  
 A/Cross-references: GDB:222816; OMIM:600494  
 A/Map position: 6q16-6q16  
 C/Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology  
 C/Keywords: alternative initiators; DNA binding; homeobox; nucleus; transcription regula  
 F:1-443/Product: transcription factor Brn-2 #status experimental <MAT1>  
 F:68-90/Region: glycine-rich  
 F:125-149/Region: histidine/proline-rich  
 F:151-165/Region: histidine/proline-rich  
 F:181-443/Product: transcription factor Oct-5a #status experimental <MAT2>  
 F:200-443/Product: transcription factor Oct-5b #status experimental <MAT3>  
 F:211-259/Region: histidine/proline-rich  
 F:269-336/Domain: POU domain homology <POU>  
 F:355-411/Domain: homeobox homology <HOX>

Query Match 70.9%; Score 100; DB 1; Length 443;  
 Best Local Similarity 89.5%; Pred. No. 0.00041;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 LRRGGGGGGGGGGGGGGG 24  
 |||||  
 Db 65 LSHGGGGGGGGGGGGGGG 83

## RESULT 4

S31224  
 transcription factor Brn-2 - mouse  
 N/Alternate names: class III POU domain protein brain-2  
 C/Species: Mus musculus (house mouse)  
 C/Date: 02-Dec-1993 #sequence\_revision 01-Sep-1995 #text\_change 22-Jun-1999

C/Accession: S31224  
 R/Hara, Y.; Rovescalli, A.C.; Kim, Y.; Nirenberg, M.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 3280-3284, 1992  
 A/Title: Structure and evolution of four POU domain genes expressed in mouse brain.  
 A/Reference number: S31223; MUID:92228768; PMID:1565620

A/Accession: S31224  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-445 <HAR>  
 A/Cross-references: EMBL:M88300; NID:G200446; PIDN:AAA39961.1; PID:G200447  
 C/Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology  
 C/Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F:68-90/Region: glycine-rich  
 F:125-151/Region: glutamine-rich  
 F:153-165/Region: histidine/proline-rich  
 F:213-261/Region: histidine/proline-rich  
 F:327-413/Domain: POU domain homology <POU>  
 F:357-413/Domain: homeobox homology <HOX>

Query Match 70.9%; Score 100; DB 1; Length 445;  
 Best Local Similarity 89.5%; Pred. No. 0.00041;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 LRRGGGGGGGGGGGGGGG 24  
 |||||  
 Db 65 LSHGGGGGGGGGGGGGGG 83

## RESULT 5

KNRZG1  
 glycine-rich cell wall structural protein 1 precursor (clone lambda-313) - rice  
 C/Species: Oryza sativa (rice)  
 C/Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 16-Jul-1999  
 C/Accession: S13385  
 R/Lel, M.; Wu, R.  
 Plant Mol. Biol. 16, 187-198, 1991  
 A/Title: A novel glycine-rich cell wall protein gene in rice.  
 A/Reference number: S13385; MUID:91370862; PMID:1716496  
 A/Accession: S13385  
 A/Molecule type: DNA  
 A/Residues: 1-165 <LEI>  
 A/Cross-references: EMBL:X53596; NID:G20246; PIDN:CAA37665.1; PID:G20247  
 C/Genetics:  
 A/Gene: gfp-1  
 C/Superfamily: glycine-rich cell wall structural protein 1  
 C/Keywords: cell wall; duplication; structural protein  
 F:1-23/Domain: signal sequence #status predicted <SIG>  
 F:24-165/Product: glycine-rich cell wall structural protein 1 #status predicted <MAT>  
 F:30-55/Region: repeat R1  
 F:56-62/Region: repeat R2  
 F:62-92/Region: repeat R1  
 F:93-99/Region: repeat R2  
 F:100-131/Region: repeat R1  
 F:132-136/Region: repeat R2  
 F:139-160/Region: repeat R1

Query Match 70.2%; Score 99; DB 1; Length 165;  
 Best Local Similarity 94.4%; Pred. No. 0.00023;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 7 KRGGGGGGGGGGGGGGG 24  
 |||||  
 Db 139 KGGGGGGGGGGGGGGGG 156

## RESULT 6

T49109  
 glycine-rich protein - Arabidopsis thaliana  
 N/Alternate names: protein At4g22020  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Sep-2000  
 C/Accession: T49109  
 R/Devan, M.; Wedler, H.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.; M



Submitted to the Protein Sequence Database, May 2000  
A:Reference number: 225016

A:Accession: T49109  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-396 <BEV>

A:Cross-references: EMBL:AL022140; GSPDB:GN00062; ATSP:AT4922020  
A:Experimental source: cultivar Columbia; BAC clone FIN20

C:Genetics:  
A:Gene: ATSP:AT4922020

A:Map position: 4  
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 70.2%; Score 99; DB 2; Length 396;  
Best Local Similarity 75.0%; Pred. No. 0.00048;  
Matches 18; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 FFPVHLKRGCGGGGGGGGGGGG 24  
DB 50 FYGKAKRYGGGGGGGGGGGGG 73

## RESULT 7

A:1948  
alpha-fetoprotein enhancer-binding protein - human

N:Alternate names: ATBF1 protein  
C:Species: Homo sapiens (man)

C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 15-Oct-1999  
R:Morinaga, T.; Yasuda, H.; Hashimoto, T.; Higashio, K.; Tamaki, T.

MOL. CELL. BIOL. 11, 6041-6049, 1991  
A:Title: A human alpha-fetoprotein enhancer-binding protein, ATBF1, contains four homeod

A:Reference number: A41948; MUID:92049333; PMID:1719379  
A:Accession: A41948

A:Molecule type: mRNA  
A:Residues: 1-2783 <MOR>

A:Cross-references: GB:D10250; GB:D90395; NID:g219429; PID:BA01095.1; PID:g219430  
A>Note: sequence extracted from NCBI Backbone (NCBIN:66271, NCBIP:66276)

C:Genetics:  
A:Gene: GDB:ATBF1

A:Cross-references: GDB:392090; OMIM:104155  
A:Map position: 16q22.3-16q23.1

C:Superfamily: alpha-fetoprotein enhancer-binding protein; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation; zinc finger

F:72-94/Region: zinc finger CCHH motif  
F:128-150/Region: zinc finger CCHH motif

F:176-198/Region: zinc finger CCHH motif  
F:311-332/Region: zinc finger CCHH motif

F:340-361/Region: zinc finger CCHH motif  
F:448-471/Region: zinc finger CCHH motif

F:489-509/Region: zinc finger CCHH motif  
F:517-538/Region: zinc finger CCHH motif

F:633-655/Region: zinc finger CCHH motif  
F:664-706/Region: zinc finger CCHH motif

F:719-773/Region: serine/threonine-rich  
F:809-958/Region: glutamine-rich

F:1071-1092/Region: zinc finger CCHH motif  
F:1117-1211/Region: proline-rich

F:1232-1288/Domain: homeobox homology <HOX1>  
F:1329-1385/Domain: homeobox homology <HOX2>

F:1416-1437/Region: zinc finger CCHH motif  
F:1618-1638/Region: zinc finger CCHH motif

F:1729-1784/Domain: homeobox homology <HOX3>  
F:1798-1820/Region: zinc finger CCHH motif

F:2033-2089/Domain: homeobox homology <HOX4>  
F:2112-2134/Region: zinc finger CCHH motif

F:2545-2566/Region: zinc finger CCHH motif  
F:2585-2607/Region: glycine-rich

F:2611-2633/Region: zinc finger CCHH motif  
F:2650-2737/Region: serine/threonine-rich

Query Match 69.5%; Score 98; DB 1; Length 2783;  
Best Local Similarity 76.2%; Pred. No. 0.0032;

Matches 16; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
OY 4 VHLKRGCGGGGGGGGGGGG 24  
DB 2580 LHVPTGGGGGGGGGGGGG 2600

## RESULT 8

T02745  
nucleic acid binding protein - rice

C:Species: Oryza sativa (rice)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 01-Dec-2000

C:Accession: T02745; T02718  
R:Yoon, U.H.; Hahn, J.H.; Yun, C.H.; Eun, M.Y.

submitted to the EMBL Data Library, February 1998  
A:Description: The rice genome contains at least two different genes encoding nucleic ac

A:Reference number: Z14712  
A:Accession: T02718

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA

A:Residues: 1-272 <YOO>  
A:Cross-references: EMBL:AF047428; NID:g4091116; PID:g4091117

A:Experimental source: strain ilpoombyeo  
R:Yoon, U.H.; Hahn, J.H.; Yun, C.H.; Eun, M.Y.

submitted to the EMBL Data Library, January 1998  
A:Description: Cloning and molecular characterization of nucleic acid binding protein ge

A:Reference number: Z14705  
A:Accession: T02718

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA

A:Residues: 1-181,183-272 <Y02>  
A:Cross-references: EMBL:AF045571; NID:g2854124; PID:g2854125

A:Experimental source: strain ilpoombyeo

Query Match 68.8%; Score 97; DB 2; Length 272;  
Best Local Similarity 84.2%; Pred. No. 0.00058;  
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 6 LKRGCGGGGGGGGGGGG 24  
DB 1 MEGCGGGGGGGGGGGG 19

## RESULT 9

A49447  
transcription factor Brn-2 - rat

N:Alternate names: class III POU domain protein brain-2  
C:Species: Rattus norvegicus (Norway rat)

C:Date: 07-Apr-1994 #sequence\_revision 18-Nov-1994 #text\_change 20-Feb-1998  
C:Accession: A49447

R:Li, P.; He, X.; Gerreiro, M.R.; Mok, M.; Aggarwal, A.; Rosenfeld, M.G.  
Genes Dev. 7, 2483-2496, 1993

A:Title: Spacing and orientation of bipartite DNA-binding motifs as potential functional  
A:Reference number: A49447; MUID:9410531; PMID:8276233

A:Accession: A49447  
A:Molecule type: mRNA

A:Status: preliminary; not compared with conceptual translation  
A:Residues: 1-445 <LII>

A:Cross-references: GB:L27663; NID:g443687  
A:Experimental source: brain

C:Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

F:68-90/Region: glycine-rich  
F:125-151/Region: glutamine-rich

F:153-165/Region: histidine/proline-rich  
F:213-261/Region: histidine/proline-rich

F:271-338/Domain: POU domain homology <POU>  
F:357-413/Domain: homeobox homology <HOX>

Query Match 68.8%; Score 97; DB 1; Length 445;  
Best Local Similarity 85.0%; Pred. No. 0.00088;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;



A:Description: catalyzes the hydrolysis of peptides  
 A:Note: cleaves preferentially after tyrosine, methionine, or arginine residues and before  
 C:Superfamily: calpain small chain; calmodulin repeat homology  
 C:Keywords: acetylated amino end; calcium binding; cysteine proteinase; duplication; EF  
 F:1-54/Domain: glycine-rich <GLY>  
 F:94-125/Domain: calmodulin repeat homology <EF1>  
 F:137-169/Domain: calmodulin repeat homology <EF2>  
 F:170-199/Domain: calmodulin repeat homology <EF3>  
 F:202-234/Domain: calmodulin repeat homology <EF4>  
 F:235-266/Domain: calmodulin repeat homology <EF5>  
 F:1/Modified site: acetylated amino end (Met) #statue experimental

Query Match 68.1%; Score 96; DB 1; Length 266;  
 Best Local Similarity 100.0%; Pred. No. 0.00073;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 GGGGGGGGGGGGGG 24  
 DB 37 GGGGGGGGGGGGGG 52

RESULT 15

C1HUL

calpain (EC 3.4.22.17) small chain - human  
 N:Alternate names: calcium-activated neutral proteinase (CANP)  
 C:Species: Homo sapiens (man)  
 C:Date: 28-Dec-1987 #sequence \_revision 28-Dec-1987 #text\_change 16-Jul-1999  
 C:Accession: A26107; A23650  
 R:Miyake, S.; Emori, Y.; Suzuki, K.  
 Nucleic Acids Res. 14, 8805-8817, 1986  
 A:Title: Gene organization of the small subunit of human calcium-activated neutral prote  
 A:Reference number: A93648; MUID:87066759; PMID:3024120  
 A:Accession: A26107  
 A:Molecule type: DNA  
 A:Residues: 1-268 <MTV>  
 A:Cross-references: GB:M31502  
 R:Ohno, S.; Emori, Y.; Suzuki, K.  
 Nucleic Acids Res. 14, 5559, 1986  
 A:Title: Nucleotide sequence of a cDNA coding for the small subunit of human calcium-dep  
 A:Reference number: A93631; MUID:86286563; PMID:3016651  
 A:Accession: A23650  
 A:Molecule type: mRNA  
 A:Residues: 1-268 <OHN>  
 A:Cross-references: EMBL:X04106; NID:935327; PIDN:CAA27726.1; PID:935328  
 C:Genetics:  
 A:Gene: GDB:CAPN4  
 A:Cross-references: GDB:119752; OMIM:114170  
 A:Map position: 19pter-19qter  
 A:Introns: 70/2; 81/3; 111/3; 131/1; 152/3; 175/3; 202/1; 241/1; 260/3  
 C:Complex: heterodimer of L (large) and S (small) chains  
 C:Function:  
 A:Description: catalyzes the hydrolysis of peptides  
 A:Note: cleaves preferentially after tyrosine, methionine, or arginine residues and before  
 C:Superfamily: calpain small chain; calmodulin repeat homology  
 C:Keywords: calcium binding; cysteine proteinase; duplication; EF hand; heterodimer; hyd  
 F:1-56/Domain: glycine-rich <GLY>  
 F:96-127/Domain: calmodulin repeat homology <EF1>  
 F:139-171/Domain: calmodulin repeat homology <EF2>  
 F:172-201/Domain: calmodulin repeat homology <EF3>  
 F:204-236/Domain: calmodulin repeat homology <EF4>  
 F:237-268/Domain: calmodulin repeat homology <EF5>

Query Match 68.1%; Score 96; DB 1; Length 266;  
 Best Local Similarity 100.0%; Pred. No. 0.00074;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 GGGGGGGGGGGGGG 24  
 DB 37 GGGGGGGGGGGGGG 52

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 15:43:05 ; Search time 4.2449 Seconds  
(without alignments)

265.882 Million cell updates/sec

Title: US-09-788-308d-4

Sequence: 1 FPPVHLKRGGGGGGGGGGGGGGG 24

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	101	71.6	476	1 EVX2_HUMAN	Q03828 homo sapien
2	101	71.6	520	1 WASP_MOUSE	P70315 mus musculu
3	101	71.6	688	1 BOMB_MOUSE	O54839 mus musculu
4	100	70.9	443	1 PO33_HUMAN	P20265 homo sapien
5	100	70.9	445	1 PO33_MOUSE	P31360 mus musculu
6	100	70.9	1627	1 TP2B_CHICK	O42131 gallus gall
7	99	70.2	165	1 GRP1_ORYSA	P25074 oryza sativ
8	99	70.2	282	1 TRP2_HUMAN	Q13595 homo sapien
9	99	70.2	485	1 ONC2_HUMAN	O09548 homo sapien
10	99	70.2	1093	1 PER_DROMI	O03297 drosophila
11	98	69.5	3703	1 ABF1_HUMAN	O15911 homo sapien
12	97	68.8	416	1 R23B_MOUSE	P54728 mus musculu
13	97	68.8	445	1 PO33_RAT	P56222 rattus norv
14	96	68.1	151	1 GDF7_MOUSE	P43029 mus musculu
15	96	68.1	263	1 CANS_BOVIN	P13135 bos taurus
16	96	68.1	266	1 CANS_PIG	P04574 sus scrofa
17	96	68.1	268	1 CANS_HUMAN	P04632 homo sapien
18	96	68.1	303	1 CH1B_POPTR	P29031 populus ttr
19	96	68.1	331	1 SHL_CERCA	O60902 homo sapien
20	96	68.1	348	1 SHL_CERCA	O61374 ceratilis c
21	96	68.1	384	1 GRP1_PETRY	O09789 petunia hyb
22	96	68.1	410	1 PO42_HUMAN	O12837 homo sapien
23	96	68.1	433	1 HXB3_MOUSE	P09026 mus musculu
24	96	68.1	447	1 COAT_ADVG	P24029 aleutian m
25	96	68.1	663	1 DUSE_MOUSE	O09112 mus musculu
26	96	68.1	911	1 ANDR_PANTR	O97775 pan troglod
27	96	68.1	919	1 ANDR_HUMAN	P10275 homo sapien
28	95	67.4	378	1 RUI7_MOUSE	O62376 mus musculu
29	93	66.0	199	1 AC22_TENMO	P26968 tenobrio mo
30	93	66.0	321	1 FBRL_HUMAN	P32087 homo sapien
31	92	65.2	440	1 FXGA_CHICK	O08937 gallus gall
32	92	65.2	495	1 PO33_MOUSE	P31361 mus musculu
33	92	65.2	497	1 PO33_RAT	O63262 rattus norv

34	92	65.2	500	1 PO33_HUMAN	P20264 homo sapien
35	92	65.2	3726	1 ABF1_MOUSE	O61339 mus musculu
36	91	64.5	206	1 TW51_MOUSE	P26687 mus musculu
37	91	64.5	280	1 CH1A_MAIZE	P29022 zea mays (m
38	91	64.5	404	1 CAZ_DROME	O27294 drosophila
39	90	63.8	141	1 YORE_TTV1	P19289 thermoprote
40	90	63.8	266	1 CANS_RABIT	P06813 oryctolagus
41	90	63.8	268	1 EP34_HCMVA	P16768 human cytom
42	90	63.8	321	1 PUR_MOUSE	P42669 mus musculu
43	90	63.8	322	1 PUR_HUMAN	O00577 homo sapien
44	90	63.8	331	1 SHX2_MOUSE	P70390 mus musculu
45	90	63.8	367	1 BET3_MESAU	O09029 mesocricetu

## ALIGNMENTS

```

RESULT 1
EVX2_HUMAN          STANDARD;          PRT;          476 AA.
ID EVX2_HUMAN
AC O03828;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Homeobox even-skipped homolog protein 2 (EVX-2).
GN EVX2.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Birren B., Linton L., Nuebaum C., Lander E.,
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 144-300 FROM N.A.
RX MEDLINE=91257849; PubMed=1675198;
RA D'Esposito M., Morelli F., Acampora D., Migliaccio E., Simeone A.,
RA Boncinelli E.;
DE "EVX2, a human homeobox gene homologous to the even-skipped
RT segmentation gene, is localized at the 5' end of HOXA locus on
RL chromosome 2."
CC Genomics 10:43-50(1991).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EARLY EMBRYOGENESIS AND
CC NEUROGENESIS IN A BIPHASIC MANNER.
CC -1- SIMILARITY: BELONGS TO THE EVEN-SKIPPED HOMEBOX FAMILY.
CC PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AC009336; -; NOT ANNOTATED CDS.
CC EMBL; M59983; AAA5244.1; -
CC EMBL; M59982; AAA5244.1; JOINED.
CC HSSP; P14653; 1B72.
CC Genew; HGNC:3507; EVX2.
CC MIM; 142991; -
CC GO; GO:0005634; C:nucleus; NAS.
CC InterPro; IPR001356; Homeobox.
CC Pfam; PF00046; homeobox; 1.
CC PRINTS; PR00024; HOMEBOX.
CC Prodom; PD000010; Homeobox; 1.
CC SMART; SM00389; HOX; 1.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS00071; HOMEBOX_2; 1.
CC DNA-binding; Developmental protein; Homeobox; Nuclear protein.
CC DNA_BIND 188 247 HOMEBOX.
FT

```

```

FT DOMAIN 294 301 POLY-ALA.
FT DOMAIN 304 301 POLY-ALA.
FT DOMAIN 346 351 POLY-ALA.
FT DOMAIN 356 370 POLY-ALA.
FT DOMAIN 373 378 POLY-ALA.
FT DOMAIN 398 408 POLY-ALA.
FT DOMAIN 413 434 POLY-GLY.
SQ SEQUENCE 476 AA; 47799 MW; 6AA99041BA151C3F CRC64;

Query Match 71.6%; Score 101; DB 1; Length 476;
Best Local Similarity 100.0%; Pred. No. 0.00035;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 RGGGGGGGGGGGGGGG 24
Db 412 RGGGGGGGGGGGGGGG 428

RESULT 2
WASP MOUSE STANDARD; PRT: 520 AA.
AC P70315;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Wiskott-Aldrich syndrome protein homolog (WASP).
GN WASP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC MEDLINE=96115600; PubMed=866397;
RA Derry J.M.J., Wiedemann P., Blair P., Wang Y., Kerns J.A.,
RA Lemahieu V., Godfrey V.L., Wilkinson J.E., Francke U.;
RT "The mouse homolog of the Wiskott-Aldrich syndrome protein (WASP)
RT gene is highly conserved and maps near the scurfy (sf) mutation on
RT the X chromosome."
RL Genomics 29:471-477(1995).
CC -1- FUNCTION: POSSIBLE REGULATOR OF LYMPHOCYTE AND PLATELET FUNCTION.
CC MAY BE INVOLVED IN SIGNALING PATHWAYS WITH CYTOSKELETAL
CC (BY SIMILARITY).
CC -1- DOMAIN: THE WH1 (WASP HOMOLOG 1) DOMAIN MAY BIND A PRO-RICH
CC LIGAND.
CC -1- SIMILARITY: Contains 1 CRIB domain.
CC -1- SIMILARITY: Contains 1 WH1 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U54788; AAC52556.1; -
CC MGD: MGI:105059; Was.
CC InterPro: IPR000697; EVH1.
CC InterPro: IPR000095; PAKbox/Rhobndng.
CC InterPro: IPR001960; WH1.
CC InterPro: IPR003124; WH2.
CC Pfam: PF00786; PBD; 1.
CC Pfam: PF00568; WH2; 1.
CC Pfam: PF02205; WH2; 1.
CC SMART: SM00285; PBD; 1.
CC SMART: SM00461; WH1; 1.
CC SMART: SM00246; WH2; 1.
CC PROSITE: PS50108; CRIB; 1.
CC Repeat: Phosphorylation.
FT DOMAIN 41 147 WH1.
FT DOMAIN 240 253 CRIB.

```

```

FT REPEAT 354 363 GRSGLPXP MOTIF 1.
FT REPEAT 393 402 GRSGLPXP MOTIF 2.
FT DOMAIN 162 167 POLY-PRO.
FT DOMAIN 314 321 POLY-PRO.
FT DOMAIN 324 341 POLY-GLY.
FT DOMAIN 368 373 POLY-PRO.
FT DOMAIN 376 379 POLY-PRO.
FT DOMAIN 384 390 POLY-PRO.
FT DOMAIN 397 403 POLY-PRO.
FT DOMAIN 408 424 POLY-PRO.
FT DOMAIN 503 520 ASP/GLU-RICH (ACIDIC).
FT MOD RES 293 293 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 520 AA; 54191 MW; 9C223733C59F0C8A CRC64;

Query Match 71.6%; Score 101; DB 1; Length 520;
Best Local Similarity 100.0%; Pred. No. 0.00038;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 RGGGGGGGGGGGGGGG 24
Db 323 RGGGGGGGGGGGGGGG 339

RESULT 3
EOMD MOUSE STANDARD; PRT: 688 AA.
AC OS4639; Q90YG7;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Eomesodermin homolog.
GN EOMES OR TBR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Kilmura N., Nakashima K., Ueno M., Taga T.;
RT "A novel mammalian T-box-containing gene, Tbr2, expressed in mouse
RT developing brain."
RL Brain Res. Dev. Brain Res. 115:183-193(1999).
[2]
RP SEQUENCE OF 278-457 FROM N.A.
RX MEDLINE=98163742; PubMed=9503012;
RA Wetli S., Rues A., Evans M., Nehls M.;
RT "A combined analysis of genomic and primary protein structure defines
RT the phylogenetic relationship of new members of the T-box family."
RL Genomics 48:24-33(1998).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: Contains 1 T-box domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB01037; BA83416.1; -
CC EMBL: AF013281; AAC16233.1; -
CC HSSP: P24781; IYBR.
CC MGD: MGI:1201683; Eomes.
CC InterPro: IPR001699; TF_T-box.
CC Pfam: PF00907; T-box; 1.
CC PRINTS: PR00937; TBOX.
CC SMART: SM00425; TBOX; 1.
CC PROSITE: PS01283; TBOX_1; 1.
CC PROSITE: PS01264; TBOX_2; 1.
CC PROSITE: PS50252; TBOX_3; 1.

```

QY	1	FFPVHLKRGGGGGGGGGGGGGG	23
Db	19	FYSLESARGGGGGGGGGGGGG	41
Query Match			
Best Local Similarity	73.64	Score 101, DB 1	Length 688
Matches	17	Conservative	2; Mismatches 4; Indels 0; Gaps
RESULT 4			
PO32_HUMAN			
ID	PO32_HUMAN	STANDARD;	PRT; 443 AA.
AC	P20265; Q14960; Q9UJL0;		
DT	01-FEB-1991 (Rel. 17, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	POU domain, class 3, transcription factor 2 (Nervous-system specific octamer-binding transcription factor N-Oct-3) (Brain-specific homeobox/POU domain protein 2) (Brain-2) (Brn-2 protein).		
OS	POU3F2 OR BRN2 OR OTF7 OR OCT7.		
GN	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
NP	[1]		
RA	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RX	MEDLINE=93181199; PubMed=8441633;		
RA	Schreiber E., Tobler A., Malipiero U., Schaffner W., Fontana A.;		
RT	"DNA cloning of human N-Oct3, a nervous-system specific POU domain transcription factor binding to the octamer DNA motif.";		
RL	Nucleic Acids Res. 21:253-258(1993).		
RN	[2]		
RA	SEQUENCE FROM N.A.		
RC	TISSUE=Liver;		
RX	MEDLINE=95380176; PubMed=7651733;		
RA	Angue J., Thomson F., Murphy K., Baker E., Sutherland G.R.,		
RA	Parsons F.G., Sturm R.A.;		
RT	"The brn-2 gene regulates the melanocytic phenotype and tumorigenic potential of human melanoma cells.";		
RN	Oncogene 11:691-700(1995).		
RN	[3]		
RA	SEQUENCE FROM N.A.		
RP	Palmer S.;		
RL	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.		
RN	[4]		
RA	SEQUENCE OF 280-404 FROM N.A.		
RC	TISSUE=Brain;		
RX	MEDLINE=89295573; PubMed=2739723;		
RA	He X., Treacy M.N., Simmons D.M., Ingraham H.A., Swanson L.W.,		
RA	Rosenfeld M.G.;		
RT	"Expression of a large family of POU-domain regulatory genes in mammalian brain development.";		
RL	Nature 340:35-42(1989).		
CC	-1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS PREFERENTIALLY TO THE RECOGNITION SEQUENCE WHICH CONSISTS OF TWO DISTINCT HALF-SITES ('GGCT') AND ('TATAT'), SEPARATED BY A NONCONSERVED SPACER REGION OF 0, 2, OR 3 NUCLEOTIDES. POSITIVELY REGULATES THE GENES UNDER THE CONTROL OF CORTICOTROPIN-RELEASING HORMONE (CRH) AND CRH II PROMOTERS (BY SIMILARITY).		
CC	-1- SUBCELLULAR LOCATION: Nuclear.		
CC	-1- ALTERNATIVE PRODUCTS:		
CC	Event-Alternative initiation;		
CC	Comment=3 isoforms, N-Oct 3 (shown here), N-Oct 5A and N-Oct 5B are produced by alternative initiation.		
CC	-1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE NEUROECTODERMAL		

CC	CELL LINEAGE.
CC	-1- SIMILARITY: BELONGS TO THE POU TRANSCRIPTION FACTOR FAMILY. CLASS-
CC	3 SUBFAMILY.
CC	-1- SIMILARITY: Contains 1 homeobox domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; Z11933; CAAT7990.1; -
DR	EMBL; L37868; AAB59611.1; -
DR	EMBL; AL022395; CAB37982.1; -
DR	PIR; S29334; S29334.
DR	HSSP; P14859; 1OCT.
DR	TRANSFAC; T00630; -
DR	TRANSFAC; T00632; -
DR	TRANSFAC; T00633; -
DR	Genew; HGNC:9215; POU3F2.
DR	MIM; 600494; -
DR	GO; GO:0003700; P:transcription factor activity; TAS.
DR	InterPro; IPR001356; Homeobox.
DR	InterPro; IPR00127; POU_domain.
DR	InterPro; IPR007103; POU_homo.
DR	Pfam; PF00046; homeobox; 1.
DR	Pfam; PF00157; pou; 1.
DR	PRINTS; PR00028; POUDOMAIN.
DR	ProDom; PD000010; Homeobox; 1.
DR	ProDom; PD000583; POU_domain; 1.
DR	SMART; SM00389; HOX; 1.
DR	SMART; SM00352; POU; 1.
DR	PROSITE; PS00027; HOMEBOX_1; 1.
DR	PROSITE; PS00035; POU_1; 1.
DR	PROSITE; PS00465; POU2; 1.
DR	PROSITE; PS50077; HOMEBOX_2; 1.
KW	DNA-binding; Nuclear protein; Homeobox; Transcription regulation;
KW	Activator; Alternative initiation.
FT	CHAIN 1 443
FT	FT CHAIN 181 443
FT	FT CHAIN 200 443
FT	FT INIT MET 181 181
FT	FT INIT MET 200 200
FT	FT DOMAIN 68 90
FT	FT DOMAIN 125 149
FT	FT DOMAIN 166 336
FT	FT DNA BIND 354 413
FT	FT CONFLICT 26 26
FT	FT CONFLICT 170 170
SO	SEQUENCE 443 AA; 46907 MW; D47167ID84D5FBE1 CRC64;
QY	Query Match 70.9%; Score 100; DB 1; Length 443;
Db	Beat Local Similarity 89.5%; Pred. No. 0.00042;
	Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
	6 LKRGGGGGGGGGGGGGGGG 24
	65 LSHGGGGGGGGGGGGGGG 83
RESULT 5	
PO32_MOUSE	STANDARD; PRT; 445 AA.
AC P31360;	
DT 01-JUL-1993 (Rel. 26, Created)	
DT 01-JUL-1993 (Rel. 26, Last sequence update)	
DT 28-FEB-2003 (Rel. 41, Last annotation update)	
DE POU domain, class 3, transcription factor 2 (Nervous-system specific	

DE octamer-binding transcription factor N-Oct-3 (Brain-specific  
 DE homeobox/POU domain protein 2) (Brain-2) (Brn-2 protein).  
 GN POU3F2 OR OTF7 OR BRN2 OR BRN-2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=92228768; PubMed=1565620;  
 RA Hara Y., Rovescalli C., Kim Y., Nirenberg M.;  
 RT "Structure and evolution of four POU domain genes expressed in mouse  
 brain.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:3280-3284(1992).  
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS PREFERENTIALLY TO THE  
 CC RECOGNITION SEQUENCE WHICH CONSISTS OF TWO DISTINCT HALF-SITES,  
 CC ('GCAAT') AND ('TAAT'), SEPARATED BY A NONCONSERVED SPACER REGION  
 CC OF 0, 2, OR 3 NUCLEOTIDES. POSITIVELY REGULATES THE GENES UNDER  
 CC THE CONTROL OF CORTICOTROPIN-RELEASING HORMONE (CRH) AND CRH II  
 CC PROMOTERS (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE NEUROECTODERMAL  
 CC CELLS LINEAGE.  
 CC -1- SIMILARITY: BELONGS TO THE POU TRANSCRIPTION FACTOR FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 homeobox domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M68300; AAA3961.1; -.  
 CC PIR: S31224; S31224.  
 CC DR HSSP: P14859; 10CT.  
 CC DR MGD: MGI:101895; Pou3f2.  
 CC DR InterPro: IPR001356; Homeobox.  
 CC DR InterPro: IPR000327; POU domain.  
 CC DR InterPro: IPR001103; POU\_homo.  
 CC DR Pfam: PF00046; homeobox; 1.  
 CC DR Pfam: PF00157; pou; 1.  
 CC DR PRINTS: PR00028; POUDOMAIN.  
 CC DR PRODOM: PD000010; Homeobox; 1.  
 CC DR PRODOM: PD000583; POU\_domain; 1.  
 CC DR SMART: SM00389; HOK; 1.  
 CC DR SMART: SM00352; POU; 1.  
 CC DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 CC DR PROSITE: PS00071; HOMEBOX\_2; 1.  
 CC DR PROSITE: PS00035; POU\_1; 1.  
 CC DR PROSITE: PS00465; POU\_2; 1.  
 CC KW DNA-binding; Nuclear protein; Homeobox; Transcription regulation;  
 KW Activator.  
 FT DOMAIN 68 90 POLY-GLY.  
 FT DOMAIN 125 149 POLY-GLN.  
 FT DOMAIN 268 338 POU.  
 FT DNA\_BIND 356 415 HOMEBOX.  
 SQ SEQUENCE 445 AA; 47149 MW; 1A47F10950EBCBA CRC64;  
 QY Query Match 70.9%; Score 100; DB 1; Length 445;  
 Db Best Local Similarity 89.5%; Pred. No. 0.00043;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 6 LKRGCGGGGGGGGGGGGGG 24  
 Db 65 LSHGCGGGGGGGGGGGGGG 83  
 RESULT 6  
 TP2B\_CHICK

ID TP2B CHICK STANDARD; PRT; 1627 AA.  
 AC 042131;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE DNA topoisomerase II, beta isozyme (EC 5.99.1.3).  
 GN TOP2B.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NCBI\_TaxID=9031;  
 RN  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=21346110; PubMed=11453553;  
 RA Nimi A., Suka N., Harata M., Kikuchi A., Mizuno S.;  
 RT "Co-localization of chicken DNA topoisomerase IIalpha, but not beta,  
 RT with sites of DNA replication and possible involvement of a  
 RT C-terminal region of alpha through its binding to PCNA.";  
 RL Chromosoma 110:102-114(2001).  
 CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT  
 CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II  
 CC MAKES DOUBLE-STRAND BREAKS (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining  
 CC of double-stranded DNA.  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; GENERALLY LOCATED IN THE  
 CC NUCLEOPLASM.  
 CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH  
 CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES  
 CC RELAX ONLY NEGATIVE SUPERCOILS.  
 CC -1- SIMILARITY: Belongs to the type II topoisomerase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AB007446; BAA22540.1; -.  
 CC DR HSSP: P06786; 1BGW.  
 CC DR InterPro: IPR003594; ATPbind\_Atpase.  
 CC DR InterPro: IPR003957; CBFA\_NFYB\_topls.  
 CC DR InterPro: IPR001241; DNA\_topoisomII.  
 CC DR InterPro: IPR002205; DNA\_topoisomIV.  
 CC DR Pfam: PF00204; DNA\_gyraseB; 1.  
 CC DR Pfam: PF02518; HATPase\_c; 1.  
 CC DR PRINTS: PR00615; CCAATSUBUNITA.  
 CC DR PRINTS: PR00418; TP2B\_FAMILY.  
 CC DR PRODOM: PD000742; DNA\_topoisomIV; 1.  
 CC DR SMART: SM00387; HATPase\_c; 1.  
 CC DR SMART: SM00433; TOP2c; 1.  
 CC DR SMART: SM00434; TOP2c; 1.  
 CC DR PROSITE: PS00177; TOPOISOMERASE II; 1.  
 CC KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.  
 KW NP\_BIND 187 192 ATP (POTENTIAL).  
 FT ACT\_SITE 831 831 DNA\_CLEAVAGE (BY SIMILARITY).  
 FT DOMAIN 5 23 POLY-GLY.  
 FT DOMAIN 1265 1268 POLY-LYS.  
 FT DOMAIN 1388 1391 POLY-ASP.  
 FT DOMAIN 1393 1396 POLY-ASN.  
 SQ SEQUENCE 1627 AA; 183245 MW; 8B651D10A2CD34B CRC64;  
 QY Query Match 70.9%; Score 100; DB 1; Length 1627;  
 Db Best Local Similarity 94.4%; Pred. No. 0.0013;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 7 KRGGCGGGGGGGGGGGGGG 24  
 Db 3 KSGGCGGGGGGGGGGGGGG 20

RESULT 7  
GRPI ORYSA  
ID GRPI ORYSA STANDARD; PRT; 165 AA.  
AC P25074;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE Glycine-rich cell wall structural protein 1 precursor.  
GN GRP-1.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=4530;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Indica-IR35;  
RA MEDLINE=91370862; PubMed=1716496;  
RT Lei M., Wu R.;  
RT "A novel glycine-rich cell wall protein gene in rice."  
RT Plant Mol. Biol. 16:187-198(1991).  
RT -1- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).  
RT -1- SUBCELLULAR LOCATION: Cell wall (Potential).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X53596; CAJ37665.1; -  
DR PIR: S13385; KNRZG1.  
DR Gramene; P25074; -  
KW Cell wall; Structural protein; Repeat; Signal.  
FT SIGNAL 1 23 POTENTIAL.  
FT CHAIN 24 165 GLYCINE-RICH CELL WALL STRUCTURAL  
FT PROTEIN 1.  
FT DOMAIN 31 159 GLY-RICH.  
FT REPEAT 56 62 R2 (TYR-RICH).  
FT REPEAT 93 99 R2 (TYR-RICH).  
FT REPEAT 132 138 R2 (TYR-RICH).  
SQ SEQUENCE 165 AA; 13536 MW; E36CE31C3650AC9A CRC64;  
Query Match 70.2%; Score 99; DB 1; Length 165;  
Best Local Similarity 94.4%; Pred. No. 0.00023;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Oy 7 KRGGGGGGGGGGGGGGG 24  
Db 139 KRGGGGGGGGGGGGGGG 156  
RESULT 8  
TR2A HUMAN  
ID TR2A HUMAN STANDARD; PRT; 282 AA.  
AC Q13555;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Transformer-2 protein homolog (TRA-2 alpha).  
GN TRA2A.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC MEDLINE=96392356; PubMed=8799144;  
RT Dauwalder B., Amaya-Manzanares F., Mattox W.;

RT "A human homologue of the Drosophila sex determination factor  
RT transformer-2 has conserved splicing regulatory functions."  
RT Proc. Natl. Acad. Sci. U.S.A. 93:9004-9009(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Uterus;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Krausberg R.D., Collins F.S., Wagner L., Shennan C.M., Shuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.L., Wang J., Hele F.,  
RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udell T.B., Toohyuk S., Grunici P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhaly S.J.,  
RA Bosak S.A., McKernan K.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Hyatt R.M.,  
RA Butterfield Y.S.N., Krzywiński M.I., Skalski U., Smalins D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences."  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP FUNCTION: SUBCELLULAR LOCATION, AND PHOSPHORYLATION.  
RC TISSUE=Cervical carcinoma;  
RX MEDLINE=98206475; PubMed=9546399;  
RA Tacke R., Toyama M., Ogawa S., Manley J.L.;  
RT "Human Tra2 proteins are sequence-specific activators of pre-mRNA  
RT splicing."  
RT Cell 93:139-148(1998).  
CC -1- FUNCTION: Sequence-specific RNA-binding protein which participates  
CC in the control of pre-mRNA splicing.  
CC -1- SUBUNIT: Binds to A3 enhancer proteins Srp75, Srp55, Srp40 and  
CC Srp30.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Name=Long;  
CC IsoId=Q13595-1; Sequence=Displayed;  
CC Name=Short;  
CC IsoId=Q13595-2; Sequence=VSP\_005893, VSP\_005894, VSP\_005895;  
CC -1- PTM: Phosphorylated in the RS domains.  
CC -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.  
CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U53209; AAC50658.1; -  
DR EMBL: BC017094; AAH17094.1; -  
DR HSSP: P19339; 2SXL.  
DR GK: Q13595; -  
DR MIM: 602718; -  
DR GO: GO:0005634; C:nucleus; IDA.  
DR GO: GO:0008248; F:pre-mRNA splicing factor activity; IDA.  
DR GO: GO:0006371; P:mRNA splicing; IDA.  
DR InterPro: IPR000504; RNA\_rec\_mot.  
DR Pfam: PF00076; rrm; 1.  
DR SMART: SM00360; RRM; 1.  
DR PROSITE: PS50102; RRM; 1.  
DR PROSITE: PS00030; RRM\_RNP\_1; 1.  
KW RNA-binding; mRNA splicing; mRNA processing; Alternative splicing;  
KW Nuclear protein; Phosphorylation.



```

FT DOMAIN 30 112 ARG/SER-RICH (RS1 DOMAIN).
FT DOMAIN 119 197 RNA-BINDING (RRM).
FT DOMAIN 198 225 LINKER.
FT DOMAIN 226 282 ARG/SER-RICH (RS2 DOMAIN).
FT VARSPLIC 1 101 Missing (in isoform Short).
FT VARSPLIC 214 214 /FtId=VSP_005893.
FT VARSPLIC 215 282 /FtId=VSP_005894.
FT VARSPLIC 215 282 Missing (in isoform Short).
SQ SEQUENCE 282 AA; 32688 MW; ED55ABE7BEA023FD CRC64;

Query Match 70.2%; Score 99; DB 1; Length 282;
Best Local Similarity 81.8%; Pred. No. 0.00037;
Matches 18; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

Qy 3 PVHLKRGCGGCGGCGGCGG 24
Db 212 PTH--SGGCGGCGGCGGCGG 231

RESULT 9
ONC2_HUMAN STANDARD; PRT; 485 AA.
AC 035948;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE One cut domain family member 2 (ONECUT-2 transcription factor) (OC-2).
GN ONECUT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99115605; PubMed=9915796;
RA Jacquemin P., Lannoy V., Rousseau G.G., Lemaigre F.P.;
"OC-2, a novel mammalian member of the ONECUT class of homeodomain
transcription factors whose function in liver partially overlaps with
that of hepatocyte nuclear factor-6."
RT J. Biol. Chem. 274:2665-2671(1999).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. ACTIVATES THE TRANSCRIPTION
OF A NUMBER OF LIVER GENES SUCH AS HNF3B.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Contains 1 CUT domain.
CC -1- SIMILARITY: BELONGS TO THE CUT HOMEBOX FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.1sb-sib.ch/announce/
or send an email to license@1sb-sib.ch).
CC EMBL: Y18198; CAB38253.1; -
DR TRANSFAC: T03259; -
DR Genem; HGNC:8139; ONECUT2.
DR MTM; 604694; -
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . . TAS.
DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR InterPro; IPR007108; Cut homeo.
DR InterPro; IPR003350; Homeo CUT.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF02376; CUT; 1.
DR Pfam; PF00046; homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SMO0389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; FALSE_NEG.
DR PROSITE; PS50071; HOMEBOX 2; 1.
KM Transcription regulation; Homeobox; DNA-binding; Nuclear protein;
Activator.

```

```

FT DNA_BIND 305 391 CUT.
FT DNA_BIND 407 466 HOMEBOX.
FT DOMAIN 18 37 POLY-GLY.
FT DOMAIN 62 66 POLY-PRO.
FT DOMAIN 75 82 POLY-ALA.
FT DOMAIN 152 165 POLY-HIS.
FT DOMAIN 298 303 POLY-SER.
SQ SEQUENCE 485 AA; 52482 MW; AF21E052FBE5DA1 CRC64;

Query Match 70.2%; Score 99; DB 1; Length 485;
Best Local Similarity 81.0%; Pred. No. 0.00058;
Matches 17; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 VHLKRGCGGCGGCGGCGG 24
Db 13 LHGARGGGCGGCGGCGG 33

RESULT 10
PER_DROWI STANDARD; PRT; 1093 AA.
AC 003297; 018421; 018422; P91721; P91722;
DT 01-OCT-1993 (Rel. 27, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Period circadian protein (Fragment).
GN PER.
OS Drosophila willistoni (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7260;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Various strains;
RX MEDLINE=97357421; PubMed=9214747;
RA Gleason J.M., Powell J.R.;
"Interspecific and intraspecific comparisons of the period locus in
the Drosophila willistoni sibling species."
RT Mol. Biol. Evol. 14:741-753(1997).
RN [2]
RP SEQUENCE OF 579-646 FROM N.A.
RX MEDLINE=93196482; PubMed=8450754;
RA Pelxco A.A., Campos S., Costa R.H., Kyriacou C.P.;
"Molecular evolution of a repetitive region within the per gene of
Drosophila."
RT Mol. Biol. Evol. 10:127-139(1993).
CC -1- FUNCTION: ESSENTIAL FOR BIOLOGICAL CLOCK FUNCTIONS. DETERMINES THE
PERIOD LENGTH OF CIRCADIAN AND ULTRADIAN RHYTHMS; AN INCREASE IN
PER DOSAGE LEADS TO SHORTENED CIRCADIAN RHYTHMS AND A DECREASE
LEADS TO LENGTHENED CIRCADIAN RHYTHMS. ESSENTIAL FOR THE CIRCADIAN
RHYTHMICITY OF LOCOMOTOR ACTIVITY, ECLOSION BEHAVIOR, AND FOR THE
RHYTHMIC COMPONENT OF THE MALE COURTSHIP SONG THAT ORIGINATES IN
THE THORACIC NERVOUS SYSTEM. THE BIOLOGICAL CYCLE DEPENDS ON THE
RHYTHMIC FORMATION AND NUCLEAR LOCALIZATION OF THE TIM-PER
COMPLEX. LIGHT INDUCES THE DEGRADATION OF TIM, WHICH PROMOTES
ELIMINATION OF PER. NUCLEAR ACTIVITY OF THE HETERODIMER
COORDINATIVELY REGULATES PER AND TIM TRANSCRIPTION THROUGH A
NEGATIVE FEEDBACK LOOP. BEHAVES AS A NEGATIVE ELEMENT IN CIRCADIAN
TRANSCRIPTIONAL LOOP. DOES NOT APPEAR TO BIND DNA, SUGGESTING
INDIRECT TRANSCRIPTIONAL INHIBITION (BY SIMILARITY).
CC -1- SUBUNIT: FORMS HETERODIMER WITH TIMLESS (TIM); THE COMPLEX THEN
TRANSLLOCATES INTO THE NUCLEUS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR AT SPECIFIC PERIODS OF THE DAY.
FIRST ACCUMULATES IN THE PERINUCLEAR REGION ABOUT ONE HOUR BEFORE
TRANSLLOCATION INTO THE NUCLEUS. INTERACTION WITH TIM IS REQUIRED
FOR NUCLEAR LOCALIZATION (BY SIMILARITY).
CC -1- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY. PROBABLY BY THE
DOUBLE-TIME PROTEIN (DBT). PHOSPHORYLATION COULD BE IMPLICATED IN
THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER
PER-TIM (BY SIMILARITY).
CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN ATLIXCO.
CC -1- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.

```

CC -!- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: US1055; AAB41360.1; -  
 CC EMBL: US1056; AAB41361.1; -  
 CC EMBL: US1057; AAB41362.1; -  
 CC EMBL: US1058; AAB41363.1; -  
 CC EMBL: US1059; AAB41364.1; -  
 CC EMBL: US1060; AAB41365.1; -  
 CC EMBL: US1061; AAB41366.1; -  
 CC EMBL: US1062; AAB41367.1; -  
 CC EMBL: US1063; AAB41368.1; -  
 CC EMBL: US1064; AAB41369.1; -  
 CC EMBL: US1065; AAB41370.1; -  
 CC EMBL: US1066; AAB41371.1; -  
 CC EMBL: US1067; AAB41372.1; -  
 CC EMBL: US1068; AAB41373.1; -  
 CC EMBL: US1069; AAB41374.1; -  
 CC EMBL: US1070; AAB41375.1; -  
 CC EMBL: US1071; AAB41376.1; -  
 CC EMBL: US1072; AAB41377.1; -  
 CC EMBL: L06342; AAA28765.1; -  
 CC FLYBASE: FBgn0013161; Dm11ver.  
 CC InterPro: IPR00014; PAS\_domain.  
 CC Pfam: PF00989; PAS; 2.  
 CC SMART: SM00091; PAS; 2.  
 CC PROSITE: PS50112; PAS; 2.  
 CC DR Biochemical rhythms; Repeat; Nuclear protein; Phosphorylation;  
 CC Polymorphism.  
 CC KW NON\_TER 1 1  
 CC FT DOMAIN <1 12  
 CC FT DOMAIN 139 209  
 CC FT DOMAIN 289 359  
 CC FT DOMAIN 371 411  
 CC FT DOMAIN 7 12  
 CC FT DOMAIN 618 625  
 CC FT DOMAIN 718 734  
 CC FT DOMAIN 745 748  
 CC FT DOMAIN 759 770  
 CC FT DOMAIN 885 888  
 CC FT DOMAIN 911 917  
 CC FT VARIANT 611 611  
 CC FT VARIANT 617 617  
 CC FT VARIANT 622 622  
 CC FT VARIANT 724 724  
 CC FT VARIANT 726 726  
 CC FT VARIANT 729 734  
 CC FT VARIANT 730 734  
 CC FT VARIANT 731 734  
 CC FT VARIANT 732 734  
 CC FT VARIANT 733 733  
 CC FT VARIANT 733 734  
 CC FT VARIANT 734 734  
 CC FT VARIANT 747 747  
 CC FT VARIANT 764 766  
 CC FT VARIANT 886 886  
 CC FT NON\_TER 1093 1093  
 CC SEQUENCE 1093 AA; 115896 MW; AB6DE050267EC187 CRC64;  
 CC Query Match Best Local Similarity 70.2%; Score 99; DB 1; Length 1093;  
 CC 89.5%; Pred. No. 0.0012;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 6 LKRGGGGGGGGGGGGG 24  
 DB 715 LNTGGGGGGGGGGGG 733  
 RESULT 11  
 ID ABL1\_HUMAN STANDARD; PRT; 3703 AA.  
 AC Q15911; Q15911; Q13719;  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 16-OCT-2001 (Rel. 40; Last sequence update)  
 DT 15-SEP-2003 (Rel. 42; Last annotation update)  
 DE Alpha-fetoprotein enhancer binding protein (AT motif-binding factor)  
 DE (AT-binding transcription factor 1).  
 GN ABL1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM A).  
 RC TISSUE=Lung;  
 RX MEDLINE=96070776; PubMed=7592926;  
 RA Miura Y., Tam T., Ido A., Morinaga T., Miki T., Hashimoto T.,  
 RA Tamaoki T.;  
 RT "Cloning and characterization of an ABL1 isoform that expresses in a  
 RT neuronal differentiation-dependent manner."  
 RT J. Biol. Chem. 270:26840-26848(1995).  
 RL [2]  
 RP SEQUENCE FROM N.A. (ISOFORM B).  
 RC TISSUE=Hepatoma;  
 RX MEDLINE=92049333; PubMed=1719379;  
 RA Morinaga T., Yasuda H., Higashio K., Tamaoki T.;  
 RT "A human alpha-fetoprotein enhancer-binding protein, ABL1, contains  
 RT four homeodomains and seventeen zinc fingers."  
 RT Mol. Cell. Biol. 11:6041-6049(1991).  
 RL [3]  
 RP SEQUENCE OF 1-1190 FROM N.A. (ISOFORM A).  
 RX MEDLINE=99425270; PubMed=10493829;  
 RA Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,  
 RA Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,  
 RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,  
 RA Eichler B.E., Harris P.C., Venter J.C., Adams M.D.;  
 RT "Genome duplications and other features in 12 Mb of DNA sequence from  
 RT human chromosome 16p and 16q."  
 RT Genomics 60:295-308(1999).  
 RL [4]  
 RP SEQUENCE OF 1151-3703 FROM N.A.  
 RA Kozlowski A., McQuerry Y., Horic M.;  
 RL Submitted (JUN1998) to the EMBL/Genbank/DBJ databases.  
 CC -!- FUNCTION: Transcriptional activator that binds to the AT-rich core  
 CC sequence of the enhancer element of the AFP gene.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=A;  
 CC IsoId=Q15911-1; Sequence=Displayed;  
 CC Name=B;  
 CC IsoId=Q15911-2; Sequence=VSP\_006825;  
 CC -!- SIMILARITY: Contains 4 homeobox domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L32832; AAC14462.1; -  
 CC EMBL: D10250; BAA01095.1; -

DR EMBL: AC002044; AAC1674.1; -;  
 DR EMBL: AC004943; AAC79153.1; -;  
 DR HSSP: P20263; 10CP.  
 DR TRANSFAC: T00048; -;  
 DR TRANSFAC: T01665; -;  
 DR GeneW: HGNC:777; ATBFL.  
 DR MIM: 104155; -;  
 DR GO: GO:000534; C: nucleus; TAS.  
 DR GO: GO:0003705; F: RNA polymerase II transcription factor activity; TAS.  
 DR GO: GO:0006355; F: regulation of transcription, DNA-dependent; TAS.  
 DR InterPro: IPR001356; Homeobox.  
 DR InterPro: IPR007087; Znf\_C2H2.  
 DR Pfam: PF00046; homeobox; 4.  
 DR Pfam: PF00096; zf-C2H2; 18.  
 DR ProDom: PD000010; Homeobox; 4.  
 DR PROSITE: PS00027; Homeobox\_1; 2.  
 DR PROSITE: PS00071; Homeobox\_2; 4.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 14.  
 DR PROSITE: PS00157; ZINC\_FINGER\_C2H2\_2; 9.  
 DR Transcription regulation; Activator; zinc-finger; Metal-binding;  
 KM DNA-binding; Homeobox; Nuclear protein; Repeat; Polymorphism;  
 KM Alternative splicing.  
 KW  
 FT ZN\_FING 282 305 C2H2-TYPE.  
 FT ZN\_FING 640 663 C2H2-TYPE.  
 FT ZN\_FING 671 694 C2H2-TYPE.  
 FT ZN\_FING 726 750 C2H2-TYPE.  
 FT ZN\_FING 804 828 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 945 968 C2H2-TYPE (DEGENERATE).  
 FT ZN\_FING 984 1008 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 1040 1064 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 1088 1112 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 1223 1246 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 1252 1275 C2H2-TYPE.  
 FT ZN\_FING 1360 1385 C2H2-TYPE.  
 FT ZN\_FING 1401 1423 C2H2-TYPE.  
 FT ZN\_FING 1429 1452 C2H2-TYPE.  
 FT ZN\_FING 1545 1569 C2H2-TYPE.  
 FT ZN\_FING 1596 1620 C2H2-TYPE.  
 FT ZN\_FING 1983 2006 C2H2-TYPE.  
 FT ZN\_FING 2145 2204 HOMEBOX 1.  
 FT DNA\_BIND 2242 2301 HOMEBOX 2.  
 FT ZN\_FING 2328 2351 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 2530 2552 HOMEBOX 3.  
 FT DNA\_BIND 2641 2700 C2H2-TYPE.  
 FT ZN\_FING 2711 2734 HOMEBOX 4.  
 FT DNA\_BIND 2944 3003 C2H2-TYPE.  
 FT ZN\_FING 3024 3048 C2H2-TYPE.  
 FT ZN\_FING 3529 3553 C2H2-TYPE.  
 FT DOMAIN 104 107 POLY-PRO.  
 FT DOMAIN 460 489 POLY-GLU.  
 FT DOMAIN 770 784 POLY-ALA.  
 FT DOMAIN 1723 1743 POLY-GLN.  
 FT DOMAIN 1789 1794 POLY-GLN.  
 FT DOMAIN 1852 1857 POLY-GLN.  
 FT DOMAIN 2037 2052 POLY-PRO.  
 FT DOMAIN 3197 3209 POLY-GLN.  
 FT DOMAIN 3210 3214 POLY-PRO.  
 FT DOMAIN 3227 3231 POLY-GLN.  
 FT DOMAIN 3376 3389 POLY-GLN.  
 FT DOMAIN 3392 3395 POLY-GLN.  
 FT DOMAIN 3507 3527 POLY-GLY.  
 FT DOMAIN 3597 3600 POLY-PRO.  
 FT DOMAIN 3636 3639 POLY-SER.  
 FT VARSPIC 1 914 Missing (in isoform B).  
 FT VARIANT 3374 3374 /FTId=VAR\_011694.  
 FT VARIANT 3377 3384 /FTId=VAR\_011695.  
 FT VARIANT 3527 3527 /FTId=VAR\_011696.  
 FT CONFLICT 72 72 A -> S (IN REF. 3).  
 FT CONFLICT 422 422 A -> P (IN REF. 3).

FT CONFLICT 579 579 T -> A (IN REF. 3).  
 FT CONFLICT 767 767 I -> S (IN REF. 3).  
 FT CONFLICT 777 777 A -> V (IN REF. 3).  
 FT CONFLICT 846 849 HHRV -> RHLG (IN REF. 3).  
 FT CONFLICT 997 997 A -> S (IN REF. 3).  
 FT CONFLICT 1150 1190 EEAIEDVEGPSPEAPAEELAKDQEGASSQAKEULTDSP  
 FT CONFLICT 1150 1190 -> GEMSHRGRLGLGVALLLETSGGLFEGGVTPACPH  
 FT VFPY (IN REF. 3).  
 SQ SEQUENCE 3703 AA; 404468 MW; 0F62AF37D4DDEF856 CRC64;  
 Query Match 69.5%; Score 98; DB 1; Length 3703;  
 Best Local Similarity 76.2%; Pred. No. 0.0041;  
 Matches 16; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 4 VHLKRGGGGGGGGGGGGGG 24  
 Db 3502 LHVPTGGGGGGGGGGGGGGG 3522  
 RESULT 12  
 R23B\_MOUSE STANDARD; PRT; 416 AA.  
 AC P54728;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE UV excision repair protein RAD23 homolog B (MHR23B) (XP-C repair  
 DE complementing complex 58 kDa protein) (P58).  
 GN RAD23B OR MHR23B.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Testis;  
 RX MEDLINE=96403997; PubMed=8808275;  
 RA van der Spek P.J., Viesser C.E., Hanaoka F., Smit B.,  
 RA Hagemeijer A., Bootsma D., Hoeijmakers J.H.J.;  
 RT "Cloning, comparative mapping, and RNA expression of the mouse  
 RT homologues of the Saccharomyces cerevisiae nucleotide excision repair  
 RT gene RAD23.";  
 RL Genomics 31:20-27(1996).  
 CC -1- FUNCTION: INVOLVED IN DNA EXCISION REPAIR. MAY PLAY A PART IN DNA  
 CC -1- DAMAGE RECOGNITION AND/OR IN ALTERING CHROMATIN STRUCTURE TO  
 CC -1- ALLOW ACCESS BY DAMAGE-PROCESSING ENZYMES.  
 CC -1- SUBUNIT: HETERODIMER OF A 125 KDa SUBUNIT (P125) AND OF A  
 CC -1- 58 KDa SUBUNIT (P58).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- SIMILARITY: Contains 1 ubiquitin-like domain.  
 CC -1- SIMILARITY: Contains 2 UBA domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X92411; CAA63146.1; -;  
 DR HSSP: P54725; IDV0.  
 DR MGD: MGI:105128; Rad23b.  
 DR InterPro: IPR004806; Rad23.  
 DR InterPro: IPR006636; STI1.  
 DR InterPro: IPR000449; UBA\_domain.  
 DR InterPro: IPR000626; Ubiquitin.  
 DR Pfam: PF00627; UBA\_2.  
 DR Pfam: PF00240; ubiquitin; 1.  
 DR SMART: SM00727; Sti1; 1.  
 DR SMART: SM00165; UBA\_2.  
 DR SMART: SM00213; UBQ; 1.  
 DR TRAFs: TIGR00601; rad23; 1.

DR PROSITE; PS00030; UBA; 2.  
DR PROSITE; PS00053; UBIQUITIN 2; 1.  
KW DNA damage; DNA repair; Nuclear protein; Repeat.  
FT DOMAIN 1 79 UBIQUITIN-LIKE.  
FT DOMAIN 188 228 UBA 1.  
FT DOMAIN 371 411 UBA 2.  
FT DOMAIN 255 261 POLY-ALA.  
FT DOMAIN 262 270 POLY-THR.  
FT DOMAIN 336 355 POLY-GLY.  
SQ SEQUENCE 416 AA; 43516 MW; 13B0245A6D892205 CRC64;  
  
Query Match 68.8%; Score 97; DB 1; Length 416;  
Best Local Similarity 94.1%; Pred. No. 0.00083;  
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
OY 8 RGGGGGGGGGGGGGGGGG 24  
DB 338 QGGGGGGGGGGGGGGGGG 354  
  
RESULT 13  
ID PO32 RAT STANDARD; PRT; 445 AA.  
AC P56222;  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DE 28-PEB-2003 (Rel. 41, Last annotation update)  
POU domain, class 3, transcription factor 2 (Nervous-system specific octamer-binding transcription factor N-Oct-3) (Brain-specific homeobox/POU domain protein 2) (Brain-2 protein).  
GN POU3F2 OR OTF7 OR BRN2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
NCBI\_Taxid=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=94102531; PubMed=8276233;  
RA Li P., He X., Guerrero M.R., Mok M., Agarwal A., Rosenfeld M.G.;  
RT "Spacing and orientation of bipartite DNA-binding motifs as potential functional determinants for POU domain factors";  
RL Genes Dev. 7:2483-2496(1993).  
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS PREFERENTIALLY TO THE RECOGNITION SEQUENCE WHICH CONSISTS OF TWO DISTINCT HALF-SITES ('GCAT') AND ('TAAT'), SEPARATED BY A NONCONSERVED SPACER REGION OF 0, 2, OR 3 NUCLEOTIDES. POSITIVELY REGULATES THE GENES UNDER THE CONTROL OF CORTICOTROPIN-RELEASING HORMONE (CRH) AND CRH II PROMOTERS.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY AT HIGH LEVELS IN THE BRAIN.  
CC -1- SIMILARITY: BELONGS TO THE POU TRANSCRIPTION FACTOR FAMILY.  
CC -1- SIMILARITY: SUBFAMILY.  
CC CLASS=3 SUBFAMILY.  
CC -1- SIMILARITY: Contains 1 homeobox domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L27663; -; NOT\_ANNOTATED\_CDS.  
DR HSSP; P14859; 10CT.  
DR InterPro; IPR001356; Homeobox.  
DR InterPro; IPR000327; POU domain.  
DR InterPro; IPR007103; POU homeo.  
DR Pfam; PF00046; homeobox; 1.  
DR Pfam; PF00157; pou; 1.  
DR PRINTS; PR00028; POUDOMAIN.  
DR PRODOM; PD000010; Homeobox; 1.

DR PRODOM; PD000583; POU domain; 1.  
DR SMART; SM00389; HOX; 1.  
DR SMART; SM00352; POU; 1.  
DR PROSITE; PS00027; HOMEBOX 1; 1.  
DR PROSITE; PS00071; HOMEBOX 2; 1.  
DR PROSITE; PS00035; POU 1; 1.  
DR PROSITE; PS00465; POU 2; 1.  
KW DNA-binding; Nuclear protein; Homeobox; Transcription regulation; Activator.  
FT DOMAIN 268 338 POU.  
FT DNA BIND 356 415 HOMEBOX.  
FT DOMAIN 68 90 POLY-GLY.  
FT DOMAIN 125 151 POLY-GLN.  
FT DOMAIN 173 177 POLY-ALA.  
FT DOMAIN 241 247 POLY-PRO.  
SQ SEQUENCE 445 AA; 47172 MW; 9D2729E34E359FD3 CRC64;  
  
Query Match 68.8%; Score 97; DB 1; Length 445;  
Best Local Similarity 85.0%; Pred. No. 0.00088;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
OY 5 HLKRGGGGGGGGGGGGGGGG 24  
DB 67 HGGSGGGGGGGGGGGGGGGG 86  
  
RESULT 14  
ID GDF7 MOUSE STANDARD; PRT; 151 AA.  
AC P43029;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
Growth/differentiation factor 7 precursor (GDF-7) (Fragment).  
GN GDF7 OR GDF-7.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c; TISSUE=Liver;  
RX MEDLINE=94195427; PubMed=8145850;  
RA Storm E.E., Huynh T.V., Copeland N.G., Jenkins N.A., Kingsley D.M., Lee S.-J.;  
RT "Limb alterations in brachypodism mice due to mutations in a new member of the TGF-beta superfamily";  
RL Nature 368:639-643(1994).  
CC -1- SUBUNIT: Homodimer; disulfide-linked (By similarity).  
CC -1- SIMILARITY: Belongs to the TGF-beta family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U08339; AAA18780.1; -.  
DR PIR; S43296; S43296.  
DR HSSP; P12643; 3BMP.  
DR MGD; MGI:95690; Gdf7.  
DR InterPro; IPR002400; GF\_cysknob.  
DR InterPro; IPR001839; TGFb.  
DR Pfam; PF00019; TGF-beta; 1.  
DR PRINTS; PR00438; GFCYSKNOT.  
DR PRODOM; PD000357; TGFb; 1.  
DR SMART; SM00204; TGFb; 1.  
DR PROSITE; PS00250; TGF-BETA\_1; 1.  
KW Growth factor; Cytokine; Glycoprotein.  
FT NON\_TER 1 5 POTENTIAL.  
FT PROPEP <1 5

```
FT CHAIN 6 151 GROWTH/DIFFERENTIATION FACTOR 7.
FT DISULFID 50 116 BY SIMILARITY.
FT DISULFID 79 148 BY SIMILARITY.
FT DISULFID 83 150 BY SIMILARITY.
FT DISULFID 115 115 INTERCHAIN (BY SIMILARITY).
FT DOMAIN 1 5 POLY-ARG.
FT DOMAIN 16 41 POLY-GLY.
SQ SEQUENCE 151 AA; 15697 MW; 0B496ACB5827759 CRC64;

Query Match 68.1%; Score 96; DB 1; Length 151;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 9 GGGGGGGGGGGGGGGG 24
Db 18 GGGGGGGGGGGGGGGG 33
```

## RESULT 15

```
CANS_BOVIN STANDARD; PRT: 263 AA.
ID P13135;
AC 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Calcium-dependent protease, small subunit (Calpain regulatory subunit)
DE (Calcium-activated neutral proteinase) (CANP).
GN CAPN1 OR CAPN4.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90008911; Pubmed=2551902;
RA McClelland P., Lash J.A., Hathaway D.R.;
RT "Identification of major autolytic cleavage sites in the regulatory
RT subunit of vascular calpain II. A comparison of partial
RT amino-terminal sequences to deduced sequence from complementary
RT DNA."
RL J. Biol. Chem. 264:17428-17431(1989).
CC -1- FUNCTION: Calcium-regulated non-lysosomal thiol-protease which
CC catalyze limited proteolysis of substrates involved in
CC cytoskeletal remodeling and signal transduction.
CC -1- SUBUNIT: Heterodimer of a large (catalytic) and a small
CC (regulatory) subunit.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; Translocates to the plasma
CC membrane upon Ca++ binding (By similarity).
CC -1- SIMILARITY: Contains 4 EF-hand calcium-binding domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
```

```
FT DOMAIN 10 25 POLY-GLY.
FT DOMAIN 34 51 POLY-GLY.
SQ SEQUENCE 263 AA; 27931 MW; 9427925D5284CE1A CRC64;

Query Match 68.1%; Score 96; DB 1; Length 263;
Best Local Similarity 100.0%; Pred. No. 0.00072;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 9 GGGGGGGGGGGGGGGG 24
Db 36 GGGGGGGGGGGGGGGG 51
```

Search completed: December 3, 2003, 15:50:34  
Job time : 5.249 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 15:44:20 ; Search time 17.1429 Seconds  
(without alignments)  
361.274 Million cell updates/sec

Title: US-09-788-308d.4  
Perfect score: 141  
Sequence: 1 FPPVHLKRGSGGGGGGGGGGG 24

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	74.5	155	5	09VZK6
2	105	74.5	525	10	09AUM0
3	102.5	72.7	443	5	08IH14
4	102.5	72.7	485	5	076876
5	101	71.6	155	5	09GND8
6	101	71.6	155	5	09GP74
7	101	71.6	155	5	09GNB7
8	101	71.6	155	5	09GP73
9	101	71.6	157	5	09GP77
10	101	71.6	158	5	09SUM4
11	101	71.6	159	5	09SUM1
12	101	71.6	159	5	09SUM5
13	101	71.6	161	5	09SUM3
14	101	71.6	161	5	09SUM6
15	101	71.6	162	5	09SUM5
16	101	71.6	163	5	09SUM9

17	101	71.6	163	5	09SUM7	09SUM7 drosophila
18	101	71.6	163	5	09SUM4	09SUM4 drosophila
19	101	71.6	163	5	09SUM6	09SUM6 drosophila
20	101	71.6	164	5	09SNP2	09SNP2 drosophila
21	101	71.6	164	5	09SUM3	09SUM3 drosophila
22	101	71.6	165	5	09SNR6	09SNR6 drosophila
23	101	71.6	165	5	09SUM1	09SUM1 drosophila
24	101	71.6	165	5	09GP44	09GP44 drosophila
25	101	71.6	165	5	09SUM2	09SUM2 drosophila
26	101	71.6	165	5	09SUM2	09SUM2 drosophila
27	101	71.6	166	5	09SUM0	09SUM0 drosophila
28	101	71.6	168	5	09SUM8	09SUM8 drosophila
29	101	71.6	175	10	09LSN6	09LSN6 aradidopsis
30	101	71.6	221	10	065514	065514 aradidopsis
31	101	71.6	304	4	09NMB9	09NMB9 homo sapien
32	101	71.6	459	4	08TB25	08TB25 homo sapien
33	101	71.6	459	4	08TB21	08TB21 homo sapien
34	101	71.6	520	11	061078	061078 mus musculus
35	101	71.6	688	11	09JUL1	09JUL1 mus musculus
36	101	71.6	697	5	09GRW7	09GRW7 drosophila
37	101	71.6	698	5	09GRX4	09GRX4 drosophila
38	101	71.6	707	11	08BN22	08BN22 mus musculus
39	101	71.6	1164	10	094UB5	094UB5 oryza sativ
40	100	70.9	264	13	09DFB6	09DFB6 gallus gall
41	100	70.9	447	13	073628	073628 anolis caro
42	99	70.2	132	10	0943G4	0943G4 oryza sativ
43	99	70.2	344	13	042403	042403 gallus gall
44	99	70.2	396	10	065450	065450 aradidopsis
45	99	70.2	612	5	08T518	08T518 anopheles g

## ALIGNMENTS

RESULT 1	ID	Q9VZK6	PRELIMINARY;	PRT;	155 AA.
AC	Q9VZK6				
DT	01-MAY-2000	(T-EMBLrel. 13, Created)			
DT	01-MAY-2000	(T-EMBLrel. 13, Last sequence update)			
DT	01-OCT-2002	(T-EMBLrel. 22, Last annotation update)			
DE	CG10853 protein (LP09837).				
GN	CG10853 OR BCDNA:LP09837.				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxId=7227;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BERKELEY;				
RX	MEDLINE=20196006; PubMed=10731132;				
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,				
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,				
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,				
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,				
RA	Brandon R.C., Rogers Y.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,				
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,				
RA	Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,				
RA	Bailly R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,				
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,				
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,				
RA	Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,				
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,				
RA	de Paulis B., Delcher A., Deng Z., Mayhew D., Dew I., Dietz S.M.,				
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,				
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,				
RA	Foster C., Gabrielian A.E., Garg N.S., Galbraith W.M., Glasser K.,				
RA	Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,				
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,				
RA	Hoskin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,				
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,				
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,				

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mostreli A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Murthy D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuske D.R., Pachb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RA "The genome sequence of *Drosophila melanogaster*."  
RL Science 287:2185-2195(2000).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY.  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,  
RA Miranda A., Mungall C.J., Nunoo J., Pachb J., Paragas V., Park S.,  
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
RA Ceiniker S.,  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AE003479; AAF47815.1; -  
DR EMBL: AY075437; AAL68252.1; -  
DR FlyBase: FBgn0035478; CG10853.  
SQ SEQUENCE 155 AA; 14855 MW; EF7D78EDD16675BF CRC64;  
Query Match 74.5%; Score 105; DB 5; Length 155;  
Best Local Similarity 81.8%; Pred. No. 1.4e-05;  
Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
OY 3 PVLKRGCGGCGGCGGCGGCGG 24  
Db 49 PVLKRGCGGCGGCGGCGGCGG 70  
RESULT 2  
O9AUM0 PRELIMINARY; PRT; 525 AA.  
AC O9AUM0.  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Putative ankyrin.  
GN OSJNB0067E01.14.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OC NCBI\_TaxID=4530;  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Buehl C.R., Tian Q., Ouyang S., Moffat K.S., Hill J.N., Burr P.C.,  
RA Buehl C.R., Zismann V., Pat G., Bowman C.L., Fujii C.Y., Vanaken S.E.,  
RA Bowman C.L., Craven B., Uteideberg T.R., Khalik H., Feldblyum T.V.,  
RA Quackenbush J., White O., Salzberg S.L., Frazer C.M.,  
RA "Oryza sativa chromosome 3 BAC OSJNB0067E01 genomic sequence."  
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
RL EMBL: AC084406; AAK26128.1; -  
DR HSSP: P53041; IAT1.  
DR Gramene; O9AUM0; -  
DR InterPro; IPR002110; ANK.  
DR InterPro; IPR001440; TPR.  
DR Pfam; PF000023; ank; 4.  
DR Pfam; PF000515; TPR; 2.  
DR SMART; SM00248; ANK; 5.  
DR PROSITE; PSS0088; ANK\_REPEAT; 3.

DR PROSITE; PSS0297; ANK\_REPEAT; 1.  
KW ANK repeat; Repeat.  
SQ SEQUENCE 525 AA; 55543 MW; 637AD93B251045F4 CRC64;  
Query Match 74.5%; Score 105; DB 10; Length 525;  
Best Local Similarity 90.0%; Pred. No. 4.4e-05;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 5 HLKRGCGGCGGCGGCGGCGG 24  
Db 9 HLKRGCGGCGGCGGCGGCGG 28  
RESULT 3  
O8IH14 PRELIMINARY; PRT; 443 AA.  
AC O8IH14.  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE L021345P.  
GN EG:132E8.1.  
OS *Drosophila melanogaster* (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OC NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=y.  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,  
RA Miranda A., Mungall C.J., Nunoo J., Pachb J., Paragas V., Park S.,  
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
RA Ceiniker S.,  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BT001477; AAN71232.1; -  
DR EMBL: 443 AA; 48924 MW; 4A99A5F95BD2E6F1 CRC64;  
SQ SEQUENCE 443 AA; 48924 MW; 4A99A5F95BD2E6F1 CRC64;  
Query Match 72.7%; Score 102.5; DB 5; Length 443;  
Best Local Similarity 81.8%; Pred. No. 7.5e-05;  
Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;  
OY 3 PVLKRGCGGCGGCGGCGGCGG 24  
Db 395 PVLKRGCGGCGGCGGCGGCGG 413  
RESULT 4  
O76876 PRELIMINARY; PRT; 485 AA.  
AC O76876.  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE EG:132E8.1 protein.  
GN EG:132E8.1 OR CG3056.  
OS *Drosophila melanogaster* (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OC NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY.  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,  
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Abayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 BA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Fabios B., Delcher A., Deng Z., Mayhew A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Horita D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimbel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,  
 RA Laevo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merulov G., Milshina N.V., Mobarry C., Morris J., Mostreli A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";  
 RA Science 287:2185-2195 (2000).  
 RL (2)  
 RN  
 RP  
 RA Papatianakia G., Spanos L., Cox S., Siden-Kiamos I., Louis C.;  
 RA "Sequencing the distal X chromosome of *Drosophila melanogaster*.";  
 RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RL (3)  
 RN  
 RP  
 RA SEQUENCE FROM N.A.  
 RA Benos P.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003420; AAF4613.1; -;  
 DR EMBL; AL023893; CA19655.1; -;  
 DR HSSP; P19339; ISXL.  
 DR FLYbase; FBgn0024987; EG:132E8.1.  
 DR InterPro; IPR002343; Hsd Sx1 RNA.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 2.  
 DR PRINTS; PR00961; HDSXLRNA.  
 DR SMART; SM00360; RRM; 2.  
 DR PROSITE; PS00102; RRM; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 SQ SEQUENCE 485 AA; 53395 MW; 18CF01939A11B06E CRC64;  
 QY Query Match 72.7%; Score 102.5; DB 5; Length 485;  
 Best Local Similarity 81.8%; Pred. No. 8.2e-05;  
 Matches 18; Conservative 1; Mismatches 0; Indels 3; Gaps 1;  
 OY 3 PVHLKRGGGGGGGGGGGGGGGG 24  
 DB 395 PHL---GGGGGGGGGGGGGGGGG 413  
 RESULT 5  
 O9GND8 PRELIMINARY; PRT; 155 AA.  
 AC O9GND8;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE NONA protein (No on or off transient A) (Fragment).  
 GN NONA.  
 OS *Drosophila littoralis*.

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCB1\_TaxID=47316;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Various strains;  
 RA Huttenen S., Campesan S., Hoikkala A.;  
 RT "Intra- and interspecific nucleotide variation at the nonA gene in  
 RT *Drosophila littoralis* and *D. virilis*.";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=11;  
 RA Huttenen S., Vleita J., Hoikkala A.;  
 RT "Levels and patterns of nucleotide variability and homopolymer length  
 RT variation at the nonA gene in *Drosophila virilis* group species.";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ304365; CAC20088.1; -;  
 DR EMBL; AJ304305; CAC20028.1; -;  
 DR EMBL; AJ304308; CAC20031.1; -;  
 DR EMBL; AJ304311; CAC20034.1; -;  
 DR EMBL; AJ304317; CAC20040.1; -;  
 DR EMBL; AJ304320; CAC20043.1; -;  
 DR EMBL; AJ304323; CAC20046.1; -;  
 DR EMBL; AJ304326; CAC20049.1; -;  
 DR EMBL; AJ304329; CAC20052.1; -;  
 DR EMBL; AJ304332; CAC20055.1; -;  
 DR EMBL; AJ304335; CAC20058.1; -;  
 DR EMBL; AJ304338; CAC20061.1; -;  
 DR EMBL; AJ304341; CAC20064.1; -;  
 DR EMBL; AJ304344; CAC20067.1; -;  
 DR EMBL; AJ304353; CAC20076.1; -;  
 DR EMBL; AJ304355; CAC20079.1; -;  
 DR EMBL; AY012599; AAG48869.1; -;  
 DR FLYbase; FBgn0043410; Ditt\nona.  
 FT NON\_TER 1  
 FT NON\_TER 155  
 SQ SEQUENCE 155 AA; 15336 MW; 5DCE33592CC84657 CRC64;  
 QY Query Match 71.6%; Score 101; DB 5; Length 155;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-05;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 8 RGGGGGGGGGGGGGGGGG 24  
 DB 68 RGGGGGGGGGGGGGGGGG 84  
 RESULT 6  
 O9GP74 PRELIMINARY; PRT; 155 AA.  
 AC O9GP74;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE NONA protein (Fragment).  
 GN NONA.  
 OS *Drosophila littoralis*.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCB1\_TaxID=47316;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Alai;  
 RA Huttenen S., Campesan S., Hoikkala A.;  
 RT "Intra- and interspecific nucleotide variation at the nonA gene in  
 RT *Drosophila littoralis* and *D. virilis*.";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ304359; CAC20082.1; -;  
 DR FLYbase; FBgn0043410; Ditt\nona.  
 FT NON\_TER 1  
 FT NON\_TER 155



FT NON TER 155 155  
SQ SEQUENCE 155 AA; 15362 MW; 5DC33593769FC57 CRC64;  
Query Match 71.6%; Score 101; DB 5; Length 155;  
Best Local Similarity 100.0%; Pred. No. 4.2e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 RGGGGGGGGGGGGGGG 24  
DB 68 RGGGGGGGGGGGGGGG 84

RESULT 7  
Q9GNB7 PRELIMINARY; PRT; 156 AA.  
AC Q9GNB7; 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE NONA protein (Fragment).  
GN NONA.  
OS Drosophila littoralis.  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=47316;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Ru3, Ou4, and Sa3;  
RA Huttunen S., Campesan S., Hoikkala A.;  
RT "Intra- and interspecific nucleotide variation at the nona gene in  
RT Drosophila littoralis and D. virilis."  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ304368; CAC20091.1; -  
DR EMBL: AJ304314; CAC20037.1; -  
DR EMBL: AJ304347; CAC20070.1; -  
DR FlyBase; FBgn0043410; Dilt\nona.  
FT NON TER 1 156  
SQ SEQUENCE 156 AA; 15393 MW; 7B6202DB1A7DCD51 CRC64;

Query Match 71.6%; Score 101; DB 5; Length 156;  
Best Local Similarity 100.0%; Pred. No. 4.3e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 RGGGGGGGGGGGGGGG 24  
DB 68 RGGGGGGGGGGGGGGG 84

RESULT 8  
Q9GP73 PRELIMINARY; PRT; 156 AA.  
AC Q9GP73; 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE NONA protein (Fragment).  
GN NONA.  
OS Drosophila littoralis.  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=47316;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Ru1;  
RA Huttunen S., Campesan S., Hoikkala A.;  
RT "Intra- and interspecific nucleotide variation at the nona gene in  
RT Drosophila littoralis and D. virilis."  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ304662; CAC20085.1; -  
DR FlyBase; FBgn0043410; Dilt\nona.

FT NON TER 1 1  
SQ SEQUENCE 156 AA; 15428 MW; 808202C5D5413BF0 CRC64;  
Query Match 71.6%; Score 101; DB 5; Length 156;  
Best Local Similarity 100.0%; Pred. No. 4.3e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 RGGGGGGGGGGGGGGG 24  
DB 68 RGGGGGGGGGGGGGGG 84

RESULT 9  
Q9GP77 PRELIMINARY; PRT; 157 AA.  
AC Q9GP77; 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE NONA protein (Fragment).  
GN NONA.  
OS Drosophila littoralis.  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=47316;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sa4;  
RA Huttunen S., Campesan S., Hoikkala A.;  
RT "Intra- and interspecific nucleotide variation at the nona gene in  
RT Drosophila littoralis and D. virilis."  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ304350; CAC20073.1; -  
DR FlyBase; FBgn0043410; Dilt\nona.  
FT NON TER 1 157  
SQ SEQUENCE 157 AA; 15450 MW; 7EFAE8E984B78DA CRC64;

Query Match 71.6%; Score 101; DB 5; Length 157;  
Best Local Similarity 100.0%; Pred. No. 4.3e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 RGGGGGGGGGGGGGGG 24  
DB 68 RGGGGGGGGGGGGGGG 84

RESULT 10  
Q95UM4 PRELIMINARY; PRT; 158 AA.  
AC Q95UM4; 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE No on or off transient A (Fragment).  
GN NONA.  
OS Drosophila flavomontana (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=40367;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=0981.0;  
RA Huttunen S., Vieira J., Hoikkala A.;  
RT "Levels and patterns of nucleotide variability and homopolymer length  
RT variation at the nona gene in Drosophila virilis group species."  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY012602; AAG48872.1; -  
DR FlyBase; FBgn0046739; Dfla\nona.  
FT NON TER 1 158

FT NON\_TER 158 158  
SQ SEQUENCE 158 AA; 15791 MW; 9AC2614FC1A4C6A7 CRC64;  
Query Match 71.6%; Score 101; DB 5; Length 158;  
Best Local Similarity 100.0%; Pred. NO. 4.3e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 RCGGGGGGGGGGGGGG 24  
|||||  
DB 74 RCGGGGGGGGGGGGGG 90

RESULT 11  
Q95UM1 PRELIMINARY; PRT; 159 AA.  
ID Q95UM1  
AC Q95UM1; 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE No on or off transient A (Fragment).  
GN NONA.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=50033;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=1540;  
RA Huttunen S., Vieira J., Hoikkala A.;  
RT "Levels and patterns of nucleotide variability and homopolymer length  
variation at the nona gene in Drosophila virilis group species.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY012600; AAG48870.1;  
DR FlyBase; FBgn0046738; Dkan\|nona.  
FT NON\_TER 1 159  
FT NON\_TER 159  
SQ SEQUENCE 159 AA; 15440 MW; 207F8CC089BD8576 CRC64;

Query Match 71.6%; Score 101; DB 5; Length 159;  
Best Local Similarity 100.0%; Pred. NO. 4.4e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 RCGGGGGGGGGGGGGG 24  
|||||  
DB 64 RCGGGGGGGGGGGGGG 80

RESULT 12  
Q95UM5 PRELIMINARY; PRT; 159 AA.  
ID Q95UM5  
AC Q95UM5; 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE No on or off transient A (Fragment).  
GN NONA.  
OS Drosophila borealis (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=40368;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=1577;  
RA Huttunen S., Vieira J., Hoikkala A.;  
RT "Levels and patterns of nucleotide variability and homopolymer length  
variation at the nona gene in Drosophila virilis group species.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY012600; AAG48870.1;  
DR FlyBase; FBgn0046745; Dbor\|nona.  
FT NON\_TER 1 159  
FT NON\_TER 159

SQ SEQUENCE 159 AA; 15649 MW; 626065561C97F762 CRC64;  
Query Match 71.6%; Score 101; DB 5; Length 159;  
Best Local Similarity 100.0%; Pred. NO. 4.4e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 8 RCGGGGGGGGGGGGGG 24  
|||||  
DB 70 RCGGGGGGGGGGGGGG 86

RESULT 13  
Q95UX3 PRELIMINARY; PRT; 161 AA.  
ID Q95UX3  
AC Q95UX3; 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE No on or off transient A (Fragment).  
GN NONA.  
OS Drosophila virilis (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7244;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=B31;  
RA Huttunen S., Vieira J., Hoikkala A.;  
RT "Levels and patterns of nucleotide variability and homopolymer length  
variation at the nona gene in Drosophila virilis group species.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY012589; AAG48859.1;  
DR FlyBase; FBgn0042729; Dvir\|nona.  
FT NON\_TER 1 161  
FT NON\_TER 161  
SQ SEQUENCE 161 AA; 15593 MW; 5B1075E948C2C91B CRC64;

Query Match 71.6%; Score 101; DB 5; Length 161;  
Best Local Similarity 100.0%; Pred. NO. 4.4e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 RCGGGGGGGGGGGGGG 24  
|||||  
DB 66 RCGGGGGGGGGGGGGG 82

RESULT 14  
Q95UM6 PRELIMINARY; PRT; 161 AA.  
ID Q95UM6  
AC Q95UM6; 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE No on or off transient A (Fragment).  
GN NONA.  
OS Drosophila ezoana (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=47313;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=E20;  
RA Huttunen S., Vieira J., Hoikkala A.;  
RT "Levels and patterns of nucleotide variability and homopolymer length  
variation at the nona gene in Drosophila virilis group species.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY012600; AAG48870.1;  
DR FlyBase; FBgn0046740; Dzeo\|nona.  
DR InterPro; IPR002952; Eggshell.  
FT PRINTS; PRO1228; EGGSHELL.  
FT NON\_TER 1 161

FT NON\_TER 161 161  
SQ SEQUENCE 161 AA; 15708 MW; 28A68346D17AB98 CRC64;

Query Match 71.6%; Score 101; DB 5; Length 161;  
Best Local Similarity 100.0%; Pred. No. 4.4e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 RGGGGGGGGGGGGGGG 24  
|||||  
DB 68 RGGGGGGGGGGGGGGG 84

## RESULT 15

O9SUX5 PRELIMINARY; PRT; 162 AA.

ID O9SUX5  
AC O9SUX5; 01-DEC-2001 (TReMBLrel. 19, Created)  
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)  
DE No on or off transient A (Fragment).  
GN NONA.  
OS Drosophila virilis (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7244;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A12;  
RA Hutunen S., Vieira J., Holkkala A.;  
RT "Levels and patterns of nucleotide variability and homopolymer length  
RT variation at the nona gene in Drosophila virilis group species."  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY012584; AAC48854.1; -;  
DR FlyBase: FBgn0042729; Dvir|nona.  
FT NON\_TER 1 1  
FT NON\_TER 162 162  
SQ SEQUENCE 162 AA; 15650 MW; EFB6FEB5B4A5419 CRC64;

Query Match 71.6%; Score 101; DB 5; Length 162;  
Best Local Similarity 100.0%; Pred. No. 4.4e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 RGGGGGGGGGGGGGGG 24  
|||||  
DB 66 RGGGGGGGGGGGGGGG 82

Search completed: December 3, 2003, 15:52:35  
Job time : 18.1429 secs